

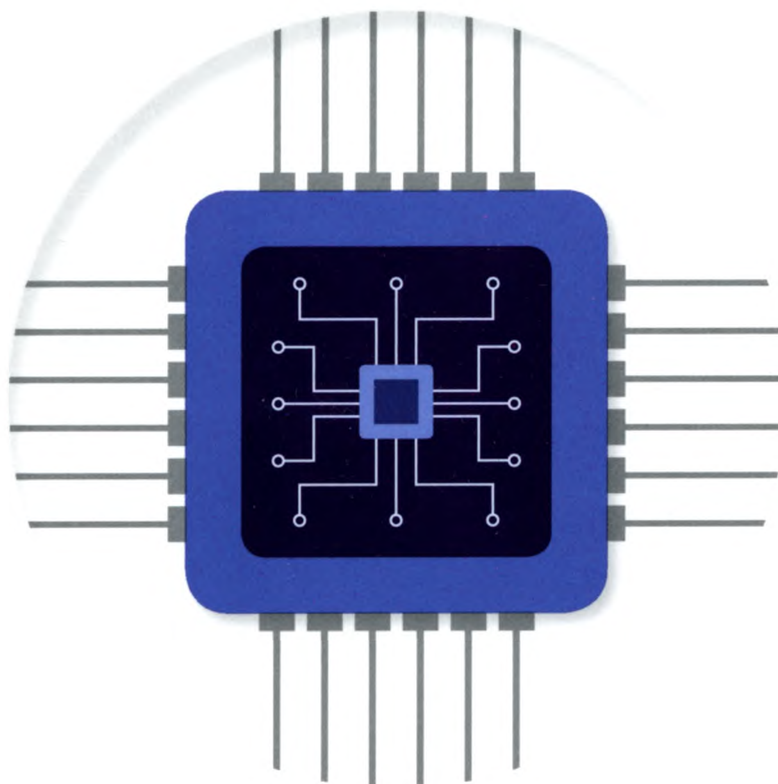
BINDER-READY EDITION: Flexible format at an even more affordable price

SEDRA/SMITH

Microelectronic Circuits

EIGHTH EDITION

ADEL S. SEDRA | KENNETH C. SMITH | TONY CHAN CARUSONE | VINCENT GAUDET



OXFORD
UNIVERSITY PRESS

Microelectronic Circuits

THE OXFORD SERIES IN ELECTRICAL AND COMPUTER ENGINEERING

Adel S. Sedra, Series Editor

Allen and Holberg, *CMOS Analog Circuit Design, 3rd edition*
Bobrow, *Elementary Linear Circuit Analysis, 2nd edition*
Bobrow, *Fundamentals of Electrical Engineering, 2nd edition*
Campbell, *Fabrication Engineering at the Micro- and Nanoscale, 4th edition*
Chen, *Digital Signal Processing*
Chen, *Linear System Theory and Design, 4th edition*
Chen, *Signals and Systems, 3rd edition*
Comer, *Digital Logic and State Machine Design, 3rd edition*
Comer, *Microprocessor-Based System Design*
Cooper and McGillem, *Probabilistic Methods of Signal and System Analysis, 3rd edition*
Dimitrijev, *Principles of Semiconductor Device, 2nd edition*
Dimitrijev, *Understanding Semiconductor Devices*
Fortney, *Principles of Electronics: Analog & Digital*
Franco, *Electric Circuits Fundamentals*
Ghausi, *Electronic Devices and Circuits: Discrete and Integrated*
Guru and Hiziroğlu, *Electric Machinery and Transformers, 3rd edition*
Houts, *Signal Analysis in Linear Systems*
Jones, *Introduction to Optical Fiber Communication Systems*
Krein, *Elements of Power Electronics, 2nd Edition*
Kuo, *Digital Control Systems, 3rd edition*
Lathi, *Linear Systems and Signals, 3rd edition*
Lathi and Ding, *Modern Digital and Analog Communication Systems, 5th edition*
Lathi, *Signal Processing and Linear Systems*
Martin, *Digital Integrated Circuit Design*
Miner, *Lines and Electromagnetic Fields for Engineers*
Mitra, *Signals and Systems*
Parhami, *Computer Architecture*
Parhami, *Computer Arithmetic, 2nd edition*
Roberts and Sedra, *SPICE, 2nd edition*
Roberts, Taenzler, and Burns, *An Introduction to Mixed-Signal IC Test and Measurement, 2nd edition*
Roulston, *An Introduction to the Physics of Semiconductor Devices*
Sadiku, *Elements of Electromagnetics, 7th edition*
Santina, Stubberud, and Hostetter, *Digital Control System Design, 2nd edition*
Sarma, *Introduction to Electrical Engineering*
Schaumann, Xiao, and Van Valkenburg, *Design of Analog Filters, 3rd edition*
Schwarz and Oldham, *Electrical Engineering: An Introduction, 2nd edition*
Sedra, Smith, Carusone, and Gaudet, *Microelectronic Circuits, 8th edition*
Stefani, Shahian, Savant, and Hostetter, *Design of Feedback Control Systems, 4th edition*
Tsividis, *Operation and Modeling of the MOS Transistor, 3rd edition*
Van Valkenburg, *Analog Filter Design*
Warner and Grung, *Semiconductor Device Electronics*
Wolovich, *Automatic Control Systems*
Yariv and Yeh, *Photonics: Optical Electronics in Modern Communications, 6th edition*
Żak, Systems and Control

EIGHTH EDITION

Microelectronic Circuits

Adel S. Sedra
University of Waterloo

Kenneth C. Smith
University of Toronto

Tony Chan Carusone
University of Toronto

Vincent Gaudet
University of Waterloo

New York Oxford
OXFORD UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trade mark of Oxford University Press in the UK and certain other countries.

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America.

For titles covered by Section 112 of the US Higher Education Opportunity Act, please visit www.oup.com/us/he for the latest information about pricing and alternate formats.

Library of Congress Cataloging-in-Publication Data

Names: Sedra, Adel S., author. | Smith, Kenneth C. (Kenneth Carless), author. | Carusone, Tony Chan, author. | Gaudet, Vincent, author.
Title: Microelectronic circuits / Adel S. Sedra, University of Waterloo, Kenneth C. Smith, University of Toronto, Tony Chan Carusone, University of Toronto, Vincent Gaudet, University of Waterloo.
Description: Eighth edition. | New York, NY: Oxford University Press, [2020] | Includes bibliographical references and index.
Identifiers: LCCN 2019017349 | ISBN 9780190853464 (acid-free paper)
Subjects: LCSH: Electronic circuits. | Microelectronics.
Classification: LCC TK7867 .S39 2020 | DDC 621.3815—dc23
LC record available at https://lccn.loc.gov/2019017349

Printing number: 9 8 7 6 5 4 3 2 1

Printed by LSC Communications, United States of America

CONTENTS IN BRIEF

CONTENTS

* Available online at www.oup.com/he/sedra-smith8e.

TABLES

FOR REFERENCE AND STUDY

For ease of reference, many of these tables are presented in the Summary Tables (ST) supplement found online at www.oup.com/he/sedra-smith8e.

HISTORICAL NOTES

PREFACE

Microelectronic Circuits, Eighth Edition, is intended as a text for the core courses in electronic circuits taught to majors in electrical and computer engineering. It should also prove useful to engineers and other professionals wishing to update their knowledge through self-study.

As was the case with the first seven editions, the objective of this book is to develop in the reader the ability to analyze and design electronic circuits, both analog and digital, discrete and integrated. While the application of integrated circuits is covered, emphasis is placed on transistor circuit design. This is done because of our belief that even if the majority of those studying this book were not to pursue a career in IC design, knowledge of what is inside the IC package would enable intelligent and innovative application of such chips. Furthermore, with the advances in VLSI technology and design methodology, IC design itself has become accessible to an increasing number of engineers.

Prerequisites

The prerequisite for studying the material in this book is a first course in circuit analysis. As a review, some linear circuits material is included here in the appendices: specifically, two-port network parameters in Appendix C; some useful network theorems in Appendix D; single-time-constant circuits in Appendix E; and s-domain analysis in Appendix F. In addition, a number of relevant circuit analysis problems are included at the beginning of the end-of-chapter problems section of Chapter 1. No prior knowledge of physical electronics is assumed. All required semiconductor device physics is included, and Appendix A provides a brief description of IC fabrication. All these appendices can be found on the book's website.

Emphasis on Design

It has been our philosophy that circuit design is best taught by pointing out the various tradeoffs available in selecting a circuit configuration and in selecting component values for a given configuration. The emphasis on design has been retained in this edition. In addition to design examples, and design-oriented exercises and end-of-chapter problems (indicated with a D), the book includes on its website an extensive appendix (Appendix B) where a large number of simulation and design examples are presented. These emphasize the use of SPICE, the most valuable circuit-design aid.

New to the Eighth Edition

The most important change in the eighth edition is that two new coauthors have joined our team: Tony Chan Carusone of the University of Toronto and Vincent Gaudet of the University of Waterloo.

While maintaining the philosophy and pedagogical approach of the first seven editions, several changes have been made to both organization and coverage. Our goal in making structural changes has been to increase modularity and thus flexibility for the instructor, without causing disturbance to courses currently using the seventh edition. Changes in coverage are necessitated by the continuing advances in technology which make some topics of greater relevance and others of less interest. As well, advances in IC process technology require that the numbers used in the examples, exercises, and end-of-chapter problems be updated to reflect the parameters of newer generations of IC technologies (e.g., some problems utilize the parameters of the 28-nm CMOS process). This ensures that students are acquiring a real-world perspective on technology.

The guiding principle in this revision has been *to make the book easier to teach and learn from*. In pursuit of this goal, the following specific and noteworthy changes have been made:

1. **New End-of-Chapter Problems.** About half of the approximately 1400 end-of-chapter problems are new or revised. To aid the instructor in deciding which of this large number of problems to assign, we have carefully selected a subset that we have designated **essential problems**. This should also be helpful to students using the book for self-study. The Instructor's Solutions Manual (ISM) has been thoroughly revised by the authors. It includes complete solutions for all exercises and end-of-chapter problems.
2. **Video Examples.** For the first time, we are including forty video examples. For each, the problem statement is provided and the student is directed to a video on the website to watch the authors solve the problem. Also, a directly related end-of-chapter problem is highlighted for the student to solve after watching the video.
3. **Summary Tables.** New and existing summary tables have been combined together and made available on the website. This collection of tables is an important resource for the student in studying and as a reference while doing homework problems.
4. **Improved Organization.** While maintaining the very successful modular organization of the seventh edition, we have reduced the number of parts of the eighth edition to three. Specifically, the filters and oscillators chapters are now in Part II: Analog Integrated Circuits.
5. **Streamlined Coverage and Book Size.** Almost every chapter has been revised and streamlined with emphasis on the essentials. This has resulted in a substantial reduction in the size of the book (by almost 200 pages). However, removed material has been made available on the website for those who want to continue to use it. Particular chapters that have been reduced are: Chapter 12 (Output Stages and Power Amplifiers); Chapter 13 (Op Amp Circuits); Chapter 14 (Filters); Chapter 15 (Oscillators); and Part III: Digital Integrated Circuits.
6. **Early Coverage of Technology Scaling and Moore's Law.** The discussion of technology scaling and Moore's law is now started in Chapter 5 (MOSFETs). It is then referenced throughout the book, and resumed in Chapter 17 (Digital Design) where the effects of scaling on the trinity of digital design—speed, power, and area—are considered.
7. **Modernizing the Study of Diodes.** Chapter 4 has been reorganized to highlight the different levels of abstraction and accuracy in diode modeling. While the coverage of standard material has been streamlined and reduced somewhat, newer topics have been expanded and/or included such as photodiodes, light-emitting diodes, application of diodes in electronic discharge (ESD) protection, etc.
8. **Clearer Derivations and Better Explanations.** Three chapters in Part II: Analog Integrated Circuits, have been thoroughly revised to simplify and clarify the presentation and to provide better derivations. These are Chapter 8 (Building Blocks of IC Amplifiers), specifically the treatment of the CG and CB amplifiers and the study of advanced current mirrors; Chapter 9 (Differential and Multistage Amplifiers), specifically the treatment of common-mode gain

and CMRR, DC offsets, and the current-mirror-loaded differential amplifier; and Chapter 10 (Frequency Response), which has been reorganized to deemphasize the study of the low-frequency response of discrete-circuit amplifiers (now placed at the end of the chapter).

9. **Clearer, Improved, and Simplified Study of Feedback.** Substantial improvements have been made to Chapter 11 (Feedback) to make the subject easier to understand and use.
10. **Streamlined and Better Organized Coverage of the Digital Topics.** Part III: Digital Integrated Circuits has undergone a thorough re-organization making it easier to integrate its topics into the first and/or the second electronics course. Its first chapter, now Chapter 16, emphasizes the basics of digital CMOS design, culminating in an in-depth study of the CMOS inverter's static characteristic. Then, Chapter 17 covers the three main metrics that are commonly used in digital circuit design and optimization, namely speed, power, and area. We then complete the discussion of technology scaling, first started in Chapter 5, by looking at how scaling impacts these three metrics. Finally, Chapter 18 focuses on transistor-level memory circuits and clocking circuits. Many of the examples, exercises, and problems in Part III have been redesigned to use newer technologies.

The Book's Website

The companion website for the book (www.oup.com/he/sedra-smith8e) contains important materials that will change frequently to reflect new developments. Here is a list of some of the materials available on the website:

1. Summary tables useful for studying and practice problems.
2. Resources to support the use of Spice with problems and examples including
   - Links to circuit simulation tools.
   - The input files needed to perform simulations of problems from the book identified with a SIM icon.
   - Additional Spice examples and the associated files.
   - Step-by-step guidance to help performing the Spice simulations.
3. Bonus text material of specialized topics that are either not covered or covered briefly in the current edition of the textbook. These include:
   - Precision Rectifier Circuits
   - Junction Field-Effect Transistors (JFETs)
   - Gallium Arsenide (GaAs) Devices and Circuits
   - Specialty Diode Topics: Diode Logic Gates, Temperature Effects in Zener Diodes, and the Schottky-Barrier Diode (SBD)
   - Useful Transistor Pairings
   - Selected Topics in BJT Ouput Stages: Class B Power Dissipation and Improvements, and Protection Circuitry
   - The Classical CMOS Class AB Configuration
   - IC Power Amplifiers
   - Power Transistor Thermal Considerations
   - The 741 Op-Amp Circuit
   - Selected Analog Filter Topics
     - First- and Second-Order Filter Functions
     - Single-Amplifier Biquadratic Active Filters
     - Sensitivity
     - Transconductance-C Filters
     - Tuned Amplifiers
   - Waveform Generators: The Monostable Multivibrator, IC Timers, and Waveform-Shaping Circuits

- MOS Velocity Saturation and Subthreshold Leakage
- Alternative Digital Logic Families
  - Pseudo-NMOS Logic Circuits
  - Dynamic MOS Logic Circuits
  - Transistor-Transistor Logic (TTL) Circuits
  - Emitter-Coupled Logic (ECL) Circuits
  - Bipolar and BiCMOS Digital Circuits
- Memory Architectures and Read-Only Memory (ROM)
- CMOS Image Sensors

4. Data sheets for hundreds of useful devices to help in laboratory experiments as well as in design projects.
5. Appendices for the Book:
   - Appendix A: VLSI Fabrication Technology
   - Appendix B: Spice Design and Simulation Examples
   - Appendix C: Two-Port Network Parameters
   - Appendix D: Some Useful Network Theorems
   - Appendix E: Single-Time-Constant Circuits
   - Appendix F: *s*-Domain Analysis: Poles, Zeros and Bode Plots
   - Appendix G: Comparison of the MOSFET and the BJT
   - Appendix H: Filter Design Tools
   - Appendix I: Bibliography
   - Appendix L: Answers to Selected Problems

Exercises and End-of-Chapter Problems

Over 450 Exercises are integrated throughout the text. The answer to each exercise is given below the exercise so students can check their understanding of the material as they read. Solving these exercises should enable the reader to gauge his or her grasp of the preceding material. In addition, more than 1400 end-of-chapter problems, half of which are new or revised in this edition, are provided. The problems are keyed to the individual chapter sections and their degree of difficulty is indicated by a rating system: difficult problems are marked with an asterisk (*); more difficult problems with two asterisks (**); and very difficult (and/or time consuming) problems with three asterisks (***). We must admit, however, that this classification is by no means exact. Our rating no doubt depended to some degree on our thinking (and mood!) at the time a particular problem was created. Answers to sample problems are given in Appendix L (on the website), so students have a checkpoint to tell if they are working out the problems correctly. Complete solutions for all exercises and problems are included in the *Instructor's Solutions Manual*, which is available from the publisher to those instructors who adopt the book.

As an aid to the instructor on deciding which to assign of this large number of problems, we have carefully selected a subset and designated it essential problems. (These are the problems with blue numbers). This should also be helpful to students using the book for self-study.

As in the previous seven editions, many examples are included. The examples, and indeed most of the problems and exercises, are based on real circuits and anticipate the applications encountered in designing real-life circuits. This edition continues the use of numbered solution steps in the figures for many examples, as an attempt to recreate the dynamics of the classroom.

Summary Tables

New and existing summary tables are presented together on the website. This collection of tables is an important resource for the student studying for exams or doing homework problems.

Video Examples

Today's students learn by watching, and they appreciate video for the ability to control the pace of presentation. For this edition, we have introduced video as a way to help students connect the text's examples to the homework problems they are assigned to solve. In 40 professionally produced videos, we walk students step by step through the procedures required to solve some of the most common, and complex, circuits they will have to master. We then provide related problems so that they can apply the strategies they have just learned to comparable circuits. We believe these videos will help students close the gap between learning and application. These videos are included in the enhanced ebook and are available to purchasers of the print book using the access code packaged with new print copies. Students with rented or used print copies can gain access to the videos by purchasing access to the ARC Premium site for *Microelectronic Circuits* at www.oup.com/he/sedra-smith8e. Videos are also available on the ARC site for instructors using *Microelectronic Circuits*.

Course Organization

The book contains sufficient material for a sequence of two single-semester courses, each of 40–50 lecture hours. The modular organization of the book provides considerable flexibility for course design. In the following, we suggest content for a sequence of two classical or standard courses. We also describe some variations on the content of these two courses and specify supplemental material for a possible third course.

The First Course

The first course is based on Part I of the book, that is, Chapters 1–7. It can be taught, most simply by starting at the beginning of Chapter 1 and concluding with the end of Chapter 7. However, as guidance to instructors who wish to follow a different order of presentation or a somewhat modified coverage, or to deal with situations where time might be constrained, we offer the following remarks:

The core of the first course is the study of the two transistor types, Chapters 5 and 6, in whatever order the instructor wishes, and transistor amplifiers in Chapter 7. These three chapters must be covered in full.

Another important part of the first course is the study of diodes (Chapter 4). Here, however, if time does not permit, some of the applications in the later part of the chapter can be skipped.

We have found it highly motivational to cover op amps (Chapter 2) near the beginning of the course. This provides the students with the opportunity to work with a practical integrated circuit and to experiment with nontrivial circuits.

Coverage of Chapter 1, at least of the amplifier sections, should prove helpful. Here the sections on signals can be either covered in class or assigned as reading material. Section 1.6 on frequency response is needed if the frequency-response of op-amp circuits is to be studied; otherwise this section can be delayed to the second course.

Finally, if the students have not taken a course on physical electronics, Chapter 3 needs to be covered. Otherwise, it can be used as review material or skipped altogether.

The Second Course

The main subject of the second course is integrated-circuit amplifiers and is based on Part II of the book, that is, Chapters 8–15. These eight chapters, however, contain more material than can be taught in one course. Thus, a judicious selection of topics to cover is called for. We hope that the following remarks can be helpful in making these choices:

The core material of Part II is presented in Chapters 8–11 and these four chapters must be covered, though not necessarily in their entirety. For instance, some of the sections near the end of a chapter and identified by the "advanced material" icon can be skipped, usually with no loss of continuity.

Beyond the required chapters (8–11), the instructor has many possibilities for the remainder of the course. These include a selection of topics from the remaining four chapters of Part II (12–15). Another possibility, is to include an introduction to digital integrated circuits by covering Chapter 16, and if time permits, selected topics of Chapters 17 and 18.

A Digitally Oriented First Course

A digitally-oriented first course can include the following: Chapter 1 (without Section 1.6), Chapter 2, Chapter 3 (if the students have not had any exposure to physical electronics), Chapter 4 (perhaps without some of the later applications sections), Chapter 5, selected topics from Chapter 7 emphasizing the basics of the application of the MOSFET as an amplifier, Chapter 16, and selected topics from Chapters 17 and 18. Such a course would be particularly suited for Computer Engineering students.

Supplemental Material/Third Course

Depending on the selection of topics for the first and second courses, some material will remain and can be used for part of a third course or as supplemental material to support student design projects. These can include Chapter 12 (Output Stages and Power Amplifiers), Chapter 13 (Op-Amp Circuits), Chapter 14 (Filters), and Chapter 15 (Oscillators), which can be used together with the advanced topics of Chapters 8–11 to support a third course on analog circuits. These can also include Chapters 16, 17, and 18 which can be used for a portion of a senior-level course on digital IC design.

The Accompanying Laboratory

Courses in electronic circuits are usually accompanied by laboratory experiments. To support the laboratory component for courses using this book, Vincent Gaudet has, in collaboration with K.C. Smith, authored a laboratory manual. *Laboratory Explorations*, together with an Instructor's Manual, is available from Oxford University Press.

An alternative approach for laboratory experimentation involves the use of pre-wired circuit boards with the experiments digitally controlled. Products that support this approach include AELabs, by Illuster Technologies, and Analog Electronic Board, by Texas Instruments; both work on the NI Elvis platform. More information can be found on the companion website (www.oup.com/he/sedra-smith8e).

An Outline for the Reader

Part I, *Devices and Basic Circuits*, includes the most fundamental and essential topics for the study of electronic circuits. At the same time, it constitutes a complete package for a first course on the subject.

Chapter 1. The book starts with an introduction to the basic concepts of electronics in Chapter 1. Signals, their frequency spectra, and their analog and digital forms are presented. Amplifiers are introduced as circuit building blocks and their various types and models are studied. This chapter also establishes some of the terminology and conventions used throughout the text.

Chapter 2. Chapter 2 deals with operational amplifiers, their terminal characteristics, simple applications, and practical limitations. We chose to discuss the op amp as a circuit building block at this early stage simply because it is easy to deal with and because the student can experiment with op-amp circuits that perform nontrivial tasks with relative ease and with a sense of accomplishment. We have found this approach to be highly motivating to the student. We should point out, however, that part or all of this chapter can be skipped and studied at a later stage (for instance, in conjunction with Chapter 9, Chapter 11, and/or Chapter 13) with no loss of continuity.

Chapter 3. Chapter 3 provides an overview of semiconductor concepts at a level sufficient for understanding the operation of diodes and transistors in later chapters. Coverage of this material is useful in particular for students who have had no prior exposure to device physics. Even those with such a background would find a review of Chapter 3 beneficial as a refresher. The instructor can choose to cover this material in class or assign it for outside reading.

Chapter 4. The first electronic device, the diode, is studied in Chapter 4. The diode terminal characteristics, the circuit models that are used to represent it, and its circuit applications are presented. Depending on the time available in the course, some of the diode applications and special diode types (Section 4.7) can be skipped or left for the student to read.

Chapters 5 and 6. The foundation of electronic circuits is established by the study of the two transistor types in use today: the MOS transistor in Chapter 5 and the bipolar transistor in Chapter 6. ***These two chapters have been written to be completely independent of one another and thus can be studied in either order, as desired.*** Furthermore, the two chapters have the same structure, making it easier and faster to study the second device, as well as to draw comparisons between the two device types.

Each of Chapters 5 and 6 begins with a study of the device structure and its physical operation, leading to a description of its terminal characteristics. Then, to allow the student to become very familiar with the operation of the transistor as a circuit element, a large number of examples are presented of dc circuits utilizing the device. The last section of each of Chapters 5 and 6 deals with second-order effects that are included for completeness, but that can be skipped if time does not permit detailed coverage. Nevertheless, we strongly recommend coverage of the newly introduced section on Moore's law and tehnology scaling in Chapter 5.

Chapter 7. The heart of a first course in electronics is the study of transistor amplifiers. Chapter 7 presents a unified treatment of the subject. It begins with the basic principles that underlie the operation of a transistor, of either type, as an amplifier, and proceeds to present the important concepts of small-signal operation and modeling. This is followed by a study of the basic configurations of single-transistor amplifiers. After a presentation of dc biasing methods, the chapter concludes with practical examples of discrete-circuit amplifiers. The combined presentation emphasizes the unity of the basic principles while allowing for separate treatment of the two device types where this is warranted. Very importantly, we are able to compare the two devices and to draw conclusions about their unique areas of application.

After the study of Part I, the reader will be fully prepared to study either analog integrated-circuits in Part II, or digital integrated circuits in Part III.

Part II, *Analog Integrated Circuits*, is devoted to the study of practical amplifier circuits that can be fabricated in the integrated-circuit (IC) form and their application in the design of filters and oscillators. Its eight chapters constitute a coherent treatment of IC amplifier design and applications and can thus serve as a second course in electronic circuits.

MOS and Bipolar. Throughout Part II, both MOS and bipolar circuits are presented side-by-side. Because the MOSFET is by far the dominant device, its circuits are presented first. Bipolar circuits are discussed to the same depth but occasionally more briefly.

Chapter 8. Beginning with a brief introduction to the philosophy of IC design, Chapter 8 presents the basic circuit building blocks that are used in the design of IC amplifiers. These include current mirrors, current sources, gain cells, and cascode amplifiers.

Chapter 9. The most important IC building block, the differential pair, is the main topic of Chapter 9. The last section of Chapter 9 is devoted to the study of multistage amplifiers.

Chapter 10. Chapter 10 presents a comprehensive treatment of the important subject of amplifier frequency response. Here, Sections 10.1 and 10.2 contain essential material; Section 10.3 provides a very useful analysis method; Sections 10.4 to 10.7 present the frequency response analysis of a variety of amplifier configurations; and Section 10.8 presents the low-frequency response of discrete-circuit amplifiers. A selection of the later sections can be made depending on the time available and the instructor's preference.

Chapter 11. The fourth of the essential topics of Part II, feedback, is the subject of Chapter 11. Both the theory of negative feedback and its application in the design of practical feedback amplifiers are presented. We also discuss the stability problem in feedback amplifiers and treat frequency compensation in some detail.

Chapter 12. In Chapter 12 we switch gears from dealing with small-signal amplifiers to those that are required to handle large signals and large amounts of power. Here we study the different amplifier classes—A, B, and AB—and their realization in bipolar and CMOS technologies. We also briefly consider power BJTs and power MOSFETs, and introduce the increasingly popular Class D amplifier. Depending on the availability of time, some of the later sections can be skipped in a first reading.

Chapter 13. Chapter 13 brings together the topics of Part II in an important application; namely, the design of operational amplifier circuits. We study both CMOS and bipolar op amps. We focus on the most fundamental circuits: the two-stage and the folded cascode op amps. We also present biasing circuits and techniques for low-voltage operation.

The last portion of Part III, Chapters 14 and 15, deals with *Filters and Oscillators*, and is intentionally oriented toward applications and systems. The two topics illustrate powerfully and dramatically the application of both negative and positive feedback.

Chapter 14. Chapter 14 deals with the design of filters, which are important building blocks of communication and instrumentation systems. A comprehensive, design-oriented treatment of the subject is presented. The material provided, together with the supplemental material in Appendix H, should allow the reader to perform a complete filter design, starting from specification and ending with a complete circuit realization. A wealth of design material is included.

Chapter 15. Chapter 15 deals with circuits for the generation of sinusoidal signals. It also includes a section on nonlinear oscillators or function generators.

Part III, *Digital Integrated Circuits*, provides a brief but nonetheless comprehensive and sufficiently detailed study of digital IC design. Our treatment is almost self-contained, requiring for the most part only a thorough understanding of the MOSFET material presented in Chapter 5. Thus Part III can be studied right after Chapter 5. The only exception to this is that knowledge of the internal capacitances of a MOSFET (Section 10.1) will be needed before taking on Chapter 17.

Chapter 16. Chapter 16 is the foundation of Part III. It begins with the motivating topic of CMOS logic-gate circuits, with a focus on switch-level implementaion of logic functions and gates. Then, following a detailed study of digital logic inverters, we concentrate on the CMOS inverter, its static characteristics, and its design. This chapter is the minimum needed to learn something meaningful about ditigal circuits.

Chapter 17. Chapter 17 presents a comprehensive overview of the so-called trinity of digital design metrics: speed, area, and power. The chapter starts by thoroughly analyzing the dynamic characteristics of a CMOS inverter. Then, transistor sizing is discussed, including the impact of sizing on speed and circuit area. Afterwards, sources of power dissipation in digital circuits are introduced. The chapter concludes by investigating the impact of semiconductor scaling—first introduced in Chapter 5—on digital circuit performance metrics.

Chapter 18. Digital circuits can be broadly divided into logic and memory circuits. The latter is the subject of Chapter 18, which first looks at the design of latches and flip-flops, and then goes into

static and dynamic cell designs for memory arrays. Finally, the chapter also introduces several useful peripheral circuits used in synchronous systems.

Appendices. The twelve appendices contain much useful background and supplementary material. We wish to draw the reader's attention in particular to the first two: Appendix A provides a concise introduction to the important topic of IC fabrication technology including IC layout. Appendix B provides SPICE device models as well as a large number of design and simulation examples in PSpice® and Multisim™. The examples are keyed to the book chapters. These Appendices and a great deal more material on these simulation examples can be found on the Companion Website.

Ancillaries

A complete set of ancillary materials is available with this text to support your course.

For the Instructor

The Ancillary Resource Center (ARC) at www.oup.com/he/sedra-smith8e is a convenient destination for all the instructor resources that accompany *Microelectronic Circuits*. Accessed online through individual user accounts, the ARC provides instructors with access to up-to-date ancillaries at any time while guaranteeing the security of grade-significant resources. On the ARC, you will find:

- **An electronic version of the Instructor's Solutions Manual.**
- **Video examples** that take students step by step through the procedures required to solve 40 problems presented in the text.
- **PowerPoint-based figure slides** that feature all the images and summary tables from the text, with their captions, so they can easily be displayed and explained in class.
- Detailed **instructor's support** for the SPICE circuit simulations.
- A set of 65 exam questions, grouped by chapter, with complete solutions, suggested time allocations, and a recommended breakdown of points per question.

The **Instructor's Solutions Manual** (ISBN 9780190853488), written by Adel Sedra, contains detailed solutions to all chapter exercises and end-of-chapter problems found in *Microelectronic Circuits*. The Instructor's Solutions Manual for ***Laboratory Explorations to Accompany Microelectronic Circuits*** (ISBN 9780197508589) contains detailed solutions to all the exercises and problems found in this student's laboratory guide; these solutions are also available online on the ARC instructor site for *Microelectronic Circuits* (www.oup.com/he/sedra-smith8e).

For the Student and Instructor

The **ARC Premium site,** available at www.oup.com/he/sedra-smith8e, features 40 professionally produced videos in which we walk students step by step through the procedures required to solve some of the most common, and complex, circuits they will have to master. ***Solved Problems*** is a set of 150 additional homework problems with complete solutions, covering concepts from the nine most used chapters in the book. This self-study aid will help students master core concepts and prepare for homework assignments and exams. Premium ARC content is included in the enhanced ebook. It is also available to purchasers of the print book using the access code packaged with new print copies. Students with rented or used print copies can purchase access codes to the ARC premium site for *Microelectronic Circuits* at www.oup.com/he/sedra-smith8e.

A Companion Website at www.oup.com/he/sedra-smith8e features permanently cached versions of device datasheets, so students can design their own circuits in class. The website also contains

SPICE circuit simulation examples and lessons. Bonus text topics and the Appendices are also featured on the website. Another very important item on the website is the Summary Tables (ST) supplement. This compilation of reference tables will benefit students completing homework assignments and studying for exams.

The ***Laboratory Explorations to Accompany Microelectronic Circuits*** (ISBN 9780197508572) invites students to explore the realm of real-world engineering through practical, hands-on experiments. Keyed to sections in the text and taking a "learn-by-doing" approach, it presents labs that focus on the development of practical engineering skills and design practices.

Acknowledgments

Many of the changes in this eighth edition were made in response to feedback received from instructors who adopted the seventh edition. We are grateful to all those who took the time to write to us. In addition, many of the reviewers provided detailed commentary on the seventh edition and suggested a number of the changes that we have incorporated in this edition. They are listed later; to all of them, we extend our sincere thanks. We are also grateful for the feedback received from the students who have taken our electronics courses over the years at the Universities of Toronto and Waterloo.

Over the recent years we have benefited greatly from discussions with a number of colleagues and friends. In particular we are very grateful to the following: James Barby, University of Waterloo; David Nairn, University of Waterloo; David Johns, University of Toronto; Ken Martin, University of Toronto; Wai-Tung Ng, University of Toronto (who wrote the original version of Appendix A and contributed to Chapter 12 in previous editions); Khoman Phang, University of Toronto; Gordon Roberts, McGill University; Ali Sheikholeslami, University of Toronto; Oliver Trescases, University of Toronto; Amir Yazdani, Ryerson University; and Derek Wright, University of Waterloo.

As she did for a number of the previous editions, Jennifer Rodrigues typed the revision with tremendous skill and good humour, and Adel Sedra is very grateful to her. Thanks also to Nijwm Wary who helped prepare the Spice ancillary material for this edition.

A large number of people at Oxford University Press contributed to the development of this edition and its various ancillaries. We would like to specifically mention Eric Sinkins of OUP Canada, who has been a tremendous support. We would also like to thank the former engineering editor, Dan Kaveney, the current engineering editor, Dan Sayre, Art Director Michele Laseau, Assistant Editor, Megan Carlson, and Production Manager, Lisa Grzan. A very special thank you goes to Senior Production Editor, Barbara Mathieu, who once more has been superb: her attention to detail and emphasis on quality is without par.

Finally, we wish to thank our families for their support and understanding, and to thank all the students and instructors who have valued this book thoroughout its history.

Adel S. Sedra
Kenneth C. Smith
Tony Chan Carusone
Vincent Gaudet
Waterloo, Ontario, Canada
September 2019

Reviewers of Eighth Edition

Marc Cahay, University of Cincinnati
Jin-Woo Choi, Louisiana State University, Baton Rouge, LA
Anthony S. Deese, The College of New Jersey, Ewing, NJ
Makarand Deo, Norfolk State University
Rohit Dua, Missouri University of Science and Technology, Springfield, MO
Nuri Emanetoglu, University of Maine, Orono, ME
Manoj Franklin, University of Maryland–College Park
Drew Hall, University of California–San Diego
Masum Hossain, University of Alberta
Yoon G. Kim, Calvin College, Grand Rapids, MI
Paul Kladitis, University of Dayton
Peter LoPresti, University of Tulsa
Edwin Marengo, Northeastern University, Boston, MA
Richard Martin, University of Delaware, Newark, DE
Bruce McNair, Stevens Institute of Technology, Hoboken, NJ
Cory Mettler, South Dakota State, Brookings, SD
Martin P. Mintchev, University of Calgary
Zoulikha Mouffak, California State University–Fresno
David Mueller, University of the Pacific, Stockton, CA
Zvi Roth, Florida Atlantic University, Boca Raton, FL
Andrew Y. J. Szeto, San Diego State University
Lili H. Tabrizi, California State University–Los Angeles
Margarita Takach, Seattle University
Ryan Christopher Toonen, University of Akron
Mehmet Vurkaç, Oregon Institute of Technology, Klamath Falls, OR
Lan Wei, University of Waterloo
Vesna Zeljkovic, Lincoln University
Dachun Zhao, California State University – Fullerton

Reviewers of Prior Editions

Maurice Aburdene, Bucknell University, Lewisburg, PA
Michael Bartz, University of Memphis, TN
Elizabeth Brauer, Northern Arizona University, Flagstaff, AZ
Martin Brooke, Duke University, Durham, NC
Jungseok Chae, Arizona State University, Tempe, AZ
Patrick L. Chapman, University of Illinois, Urbana–Champaign, IL
Yun Chiu, University of Illinois, Urbana–Champaign, IL
Roy H. Cornely, New Jersey Institute of Technology, Newark, NJ
Norman Cox, Missouri University of Science and Technology, Rolla, MO
Dale L. Critchlow, University of Vermont, Burlingon, VT
Robert Bruce Darling, University of Washington, Seattle, WA
Artice Davis, San Jose State University, CA
John Davis, University of Texas, Austin, TX
Christopher DeMarco, University of Wisconsin, Madison, WI
Lian Dong, Baylor University, Waco, TX
Robert Engelken, Arkansas State University, Jonesboro, AR
Muhammad Faisal, University of Michigan, Ann Arbor, MI
Ethan Farquhar, University of Tennessee, Knoxville, TN
Patrick Fay, University of Notre Dame, Notre Dame, IN
Eby G. Friedman, University of Rochester, NY
Paul M. Furth, New Mexico State University, Las Cruces, NM
Vincent Gaudet, University of Waterloo, Waterloo, Canada
Rhett T. George, Jr., Duke University, Durham, NC
Roobik Gharabagi, St. Louis University, MO
George Giakos, University of Akron, OH
John Gilmer, Wilkes University, Wilkes-Barre, PA
Michael Green, University of California, Irvine, CA
Elmer A. Grubbs, Northern Arizona University, Flagstaff, AZ
Steven de Haas, California State University, Sacramento, CA
Anas Hamoui, McGill University, Montreal, Canada
Reza Hashemian, Northern Illinois University, DeKalb, IL
William Harrell, Clemson University, SC
Reid Harrison, University of Utah, Salt Lake City, UT
Ward J. Helms, University of Washington, Seattle, WA
Richard Hornsey, York University, Toronto, Canada
Timothy Horiuchi, University of Maryland, College Park, MD
Hsiung Hsu, The Ohio State University, Columbus, OH
Robert Irvine, California State Polytechnic University, Pomona, CA
Mohammed Ismail, The Ohio State University, Columbus, OH
Paul Israelsen, Utah State University, Logan UT
Justin Jackson, Weber State University, Ogden, UT
Steve Jantzi, Broadcom, CA
Zhenhua Jiang, University of Miami, FL
Marian Kazimierczuk, Wright State University, Dayton, OH
John Khoury, Columbia University, New York, NY
Jacob B. Khurgin, The Johns Hopkins University, Baltimore, MD
Seongsin M. Kim, University of Alabama, Tuscaloosa, AL
Roger King, University of Toledo, OH
Clark Kinnaird, Southern Methodist University, Dallas, TX
Robert J. Krueger, University of Wisconsin, Milwaukee, WI
Joy Laskar, Georgia Institute of Technology, Atlanta, GA
Serhiy Levkov, New Jersey Institute of Technology, Newark, NJ
Tsu-Jae King Liu, University of California, Berkeley, CA
Yicheng Lu, Rutgers University, Piscataway, NJ
David Luke, University of New Brunswick, Fredericton, Canada
Leda Lunardi, North Carolina State University, Raleigh, NC
John Mankowski, Texas Tech University, Lubbock, TX
Thomas Matthews, California State University, Sacramento, CA
Chris Mi, University of Michigan, Dearborn, MI
Un-Ku Moon, Oregon State University, Corvallis, OR
Bahram Nabet, Drexel University, Philadelphia, PA
Dipankar Nagchoudhuri, Indian Institute of Technology, Delhi, India
David Nairn, University of Waterloo, Waterloo, Canada
Phyllis R. Nelson, California State Polytechnic University, Pomona, CA
Joseph H. Nevin, University of Cincinnati, OH
Robert W. Newcomb, University of Maryland, College Park, MD
Toshikazu Nishida, University of Florida, Gainesville, FL
Ken Noren, University of Idaho, Moscow, ID
Brita Olson, California Polytechnic University, Pomona, CA
Martin Peckerar, University of Maryland, College Park, MD
Khoman Phang, University of Toronto, Canada
Mahmudur Rahman, Santa Clara University, CA
Rabin Raut, Concordia University, Montreal, Canada
John A. Ringo, Washington State University, Pullman, WA
Zvi S. Roth, Florida Atlantic University, Boca Raton, FL
Mulukutla Sarma, Northeastern University, Boston, MA
John Scalzo, Louisiana State University, Baton Rouge, LA
Norman Scheinberg, City College, New York, NY
Pierre Schmidt, Florida International University, Miami, FL
Richard Schreier, Analog Devices, Toronto, Canada
Dipankar Sengupta, Royal Melbourne Institute of Technology, Australia
Ali Sheikholeslami, University of Toronto, Canada
Kuang Sheng, Rutgers University, Piscataway, NJ
Michael L. Simpson, University of Tennessee, Knoxville, TN

Karl A. Spuhl, Washington University in St. Louis, MO
Charles Sullivan, Dartmouth College, Hanover, NH
Matthew Swabey, Purdue University, West Lafayette, IN
Andrew Szeto, San Diego State University, CA
Khalid Hasan Tantawi, University of Alabama, Huntsville, AL
Joel Therrien, University of Massachusetts, Lowell, MA
Farid M. Tranjan, University of North Carolina, Charlotte, NC
Len Trombetta, University of Houston, TX
Daniel van der Weide, University of Delaware, Newark, DE
Gregory M. Wierzba, Michigan State University, East Lansing, MI
Mustapha C. E. Yagoub, University of Ottawa, Ottawa, Canada
Sandra Yost, University of Detroit, Mercy, MI
Donna Yu, North Carolina State University, Raleigh, NC
Jiann-Shiun Yuan, University of Central Florida, Orlando, FL
Alex Zaslavsky, Brown University, Providence, RI
Jianhua (David) Zhang, University of Illinois, Urbana–Champaign, IL

Microelectronic Circuits

PART I

Devices and Basic Circuits

Part I, *Devices and Basic Circuits*, includes the most fundamental and essential topics for the study of electronic circuits. At the same time, it constitutes a complete package for a first course on the subject.

The heart of Part I is the study of the three basic semiconductor devices: the diode (Chapter 4), the MOS transistor (Chapter 5), and the bipolar transistor (Chapter 6). In each case, we study the device operation, its characterization, and its basic circuit applications. Chapter 7 then follows with a study of the most fundamental application of the two transistor types; namely, their use in amplifier design. This side-by-side study of MOSFET and BJT amplifiers allows us to see similarities between these amplifiers and to compare them, which in turn highlights the distinct areas of applicability of each, as well as showing the unity of the basic principles that underlie the use of transistors as amplifiers.

For those who have not had a prior course on device physics, Chapter 3 provides an overview of semiconductor concepts at a level sufficient for the study of electronic circuits. A review of Chapter 3 should prove useful even for those with prior knowledge of semiconductors.

Since the purpose of electronic circuits is the processing of signals, it is essential to understand signals, their characterization in the time and frequency domains, and their analog and digital representations. The basis for such understanding is provided in Chapter 1, which also introduces the most common signal-processing function, *amplification*, and the characterization and types of *amplifiers*.

Besides diodes and transistors, the basic electronic devices, the op amp is studied in Part I. Although not an electronic device in the most fundamental sense, the op amp is commercially available as an integrated circuit (IC) package and has well-defined terminal characteristics. Thus, even though the op amp's internal circuit is complex, typically incorporating 20 or more transistors, its almost-ideal terminal behavior makes it possible to treat the op amp as a circuit element and to use it in the design of powerful circuits, as we do in Chapter 2, without any knowledge of its internal construction. We should mention, however, that the study of op amps can be delayed until a later point, and Chapter 2 can be skipped with no loss of continuity.

The foundation of this book, and of any electronics course, is the study of the two transistor types in use today: the MOS transistor in Chapter 5 and the bipolar transistor in Chapter 6. These two chapters have been written to be completely independent of each other and thus can be studied in either order, as desired.

After the study of Part I, the reader will be fully prepared to undertake the study of either integrated-circuit amplifiers in Part II or digital integrated circuits in Part III.

CHAPTER 1

Signals and Amplifiers

IN THIS CHAPTER YOU WILL LEARN

- That electronic circuits process signals, and thus understanding electrical signals is essential to appreciating the material in this book.
- The Thévenin and Norton representations of signal sources.
- The representation of a signal as the sum of sine waves.
- The analog and digital representations of a signal.
- The most basic and pervasive signal-processing function: signal amplification, and correspondingly, the signal amplifier.
- How amplifiers are characterized (modeled) as circuit building blocks independent of their internal circuitry.
- How the frequency response of an amplifier is measured, and how it is calculated, especially in the simple but common case of a single-time-constant (STC) type response.

Introduction

The subject of this book is modern electronics, a field that has come to be known as **microelectronics**. **Microelectronics** refers to the integrated-circuit (IC) technology that at the time of this writing is capable of producing circuits that contain billions of components in a small piece of silicon (known as a **silicon chip**) whose area is roughly 100 mm^2. One such microelectronic circuit is a complete digital computer, which is known, appropriately, as a **microcomputer** or, more generally, a **microprocessor**. The microelectronic circuits you will learn to design in this book are used in almost every device we encounter in our daily lives: in the appliances we use in our homes; in the vehicles and transportation systems we use to travel; in the cellphones we use to communicate; in the medical equipment we need to care for our health; in the computers we use to do our work; and in the audio and video systems, the gaming consoles and televisions, and the multitude of other digital devices we use to entertain ourselves. Indeed, it is difficult to conceive of modern life without microelectronic circuits.

In this book we will study electronic devices that can be used singly (in the design of **discrete circuits**) or as components of an **integrated-circuit (IC)** chip. We will study the design and analysis of interconnections of these devices, which form discrete and integrated circuits of varying complexity and perform a wide variety of functions. We will also learn about available IC chips and their application in the design of electronic systems.

The purpose of this first chapter is to introduce some basic concepts and terminology. In particular, we will learn about signals and about one of the most important signal-processing functions electronic circuits are designed to perform: signal amplification. We will then look at circuit representations or models for linear amplifiers. These models will be used in subsequent chapters in the design and analysis of actual amplifier circuits.

In addition to motivating the study of electronics, this chapter serves as a bridge between the study of linear circuits and that of the subject of this book: the design and analysis of electronic circuits. Thus, we presume a familiarity with linear circuit analysis, as in the following example.

Video Example VE 1.1

For the circuit shown in Fig. VE1.1, find the current in each of the three resistors and the voltage (with respect to ground) at their common node using two methods:

(a) Loop Equations: Define branch currents I_1 and I_2 in R_1 and R_2, respectively; write two equations and solve them.

(b) Node Equation: Define the node voltage V at the common node; write a single equation and solve it.

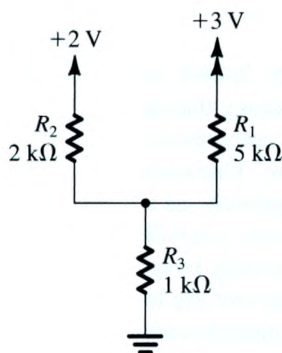


Figure VE1.1 The circuit for Video Example 1.1.


VE 1.1

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 1.17

1.1 Signals

Signals contain information about a variety of things and activities in our physical world. Examples abound: Information about the weather is contained in signals that represent the air temperature, pressure, wind speed, etc. The voice of a radio announcer reading the news into a microphone provides an acoustic signal that contains information about world affairs. To monitor the status of a nuclear reactor, instruments are used to measure a multitude of relevant parameters, each instrument producing a signal.

To extract required information from a set of signals, the observer (be it a human or a machine) invariably needs to **process** the signals in some predetermined manner. This **signal processing** is usually most conveniently performed by electronic systems. For this to be possible, however, the signal must first be converted into an electrical signal, that is, a voltage or a current. This process is accomplished by devices known as **transducers**. A variety of transducers exist, each suitable for one of the various forms of physical signals. For instance, the sound waves generated by a human can be converted into electrical signals using a microphone, which is in effect a pressure transducer. It is not our purpose here to study transducers; rather, we shall assume that the signals of interest already exist in the electrical domain and represent them by one of the two equivalent forms shown in Fig. 1.1.

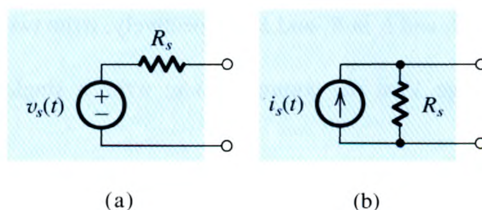


Figure 1.1 Two alternative representations of a signal source: **(a)** the Thévenin form; **(b)** the Norton form.

In Fig. 1.1(a) the signal is represented by a voltage source $v_s(t)$ with a source resistance R_s. In the alternate representation of Fig. 1.1(b) the signal is represented by a current source $i_s(t)$ with a source resistance R_s. Although the two representations are equivalent, the one in Fig. 1.1(a) (known as the Thévenin form) is preferred when R_s is low. The representation of Fig. 1.1(b) (known as the Norton form) is preferred when R_s is high. You will come to appreciate this point later in this chapter when we study the different types of amplifiers. For the time being, it is important to be familiar with Thévenin's and Norton's theorems (for a brief review, see Appendix D) and to note that for the two representations in Fig. 1.1 to be equivalent, their parameters are related by

$$v_s(t) = R_s i_s(t)$$

Example 1.1

The output resistance of a signal source, although inevitable, is an imperfection that limits the ability of the source to deliver its full signal strength to a **load**. To see this point more clearly, consider the signal source when connected to a load resistance R_L as shown in Fig. 1.2. For the case in which the source is represented by its Thévenin equivalent form, find the voltage v_o that appears across R_L, and hence the condition that R_s must satisfy for v_o to be close to the value of v_s. Repeat for the Norton-represented source, in this case finding the current i_o that flows through R_L and hence the condition that R_s must satisfy for i_o to be close to the value of i_s.

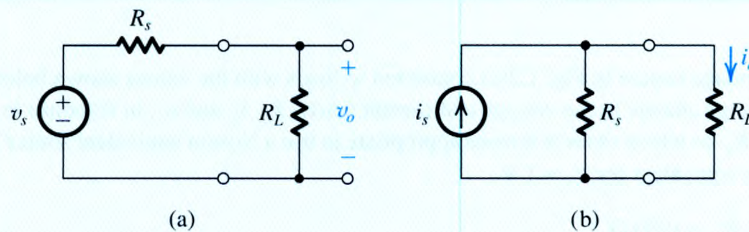


Figure 1.2 Circuits for Example 1.1.

Solution

For the Thévenin-represented signal source shown in Fig. 1.2(a), the output voltage v_o that appears across the load resistance R_L can be found from the ratio of the voltage divider formed by R_s and R_L,

$$v_o = v_s \frac{R_L}{R_L + R_s}$$

From this equation we see that as long as $R_s \ll R_L$,

$$v_o \simeq v_s$$

insensitive to small changes in R_s and R_L. Thus, for a source represented by its Thévenin equivalent, ideally $R_s = 0$, and as R_s is increased, relative to the load resistance R_L, the voltage v_o that appears across the load becomes smaller, not a desirable outcome.

Example 1.1 *continued*

Next, we consider the Norton-represented signal source in Fig. 1.2(b). To obtain the current i_o that flows through the load resistance R_L, we use the ratio of the current divider formed by R_s and R_L,

$$i_o = i_s \frac{R_s}{R_s + R_L}$$

From this relationship we see that as long as $R_s \gg R_L$,

$$i_o \simeq i_s$$

insensitive to the precise values of R_s and R_L. Thus for a signal source represented by its Norton equivalent, ideally $R_s = \infty$, and as R_s is reduced, relative to the load resistance R_L, the current i_o that flows through the load becomes smaller, not a desirable outcome.

Finally, we note that although circuit designers cannot usually do much about the value of R_s, they may have to devise a circuit solution that minimizes or eliminates the loss of signal strength that results when the source is connected to the load.

Video Example VE 1.2

Consider the voltage source in Fig. 1.2(a) connected to loads with the values shown below. In each case, find the percentage change in the voltage and current across R_L, v_o and o_o, in response to a 10% increase in the value of R_L. In which cases is it more appropriate to use a Norton equivalent source? In those cases, find the Norton equivalent for $V_s = 1$ V.

(a) $R_s = 2\ \text{k}\Omega$; $R_L = 100\ \text{k}\Omega$
(b) $R_s = 100\ \Omega$; $R_L = 8\ \Omega$
(c) $R_s = 5\ \text{k}\Omega$; $R_L = 50\ \text{k}\Omega$
(d) $R_s = 1\ \text{k}\Omega$; $R_L = 50\ \Omega$

VE 1.2

Solution: Go to **www.oup.com/he/sedra-smith8e** to watch the authors solve this problem.

Related end-of-chapter problem: 1.30

EXERCISES

1.1 For the signal-source representations shown in Figs. 1.1(a) and 1.1(b), what are the open-circuit output voltages that would be observed? If, for each, the output terminals are short-circuited (i.e., wired together), what current would flow? For the representations to be equivalent, what must the relationship be between v_s, i_s, and R_s?

Ans. For (a), $v_{oc} = v_s(t)$; for (b), $v_{oc} = R_s i_s(t)$; for (a), $i_{sc} = v_s(t)/R_s$; for (b), $i_{sc} = i_s(t)$; for equivalency, $v_s(t) = R_s i_s(t)$

1.2 A signal source has an open-circuit voltage of 10 mV and a short-circuit current of 10 μA. What is the source resistance?

Ans. 1 kΩ

1.3 A signal source that is most conveniently represented by its Thévenin equivalent has $v_s = 10$ mV and $R_s = 1$ kΩ. If the source feeds a load resistance R_L, find the voltage v_o that appears across the load for $R_L = 100$ kΩ, 10 kΩ, 1 kΩ, and 100 Ω. Also, find the lowest permissible value of R_L for which the output voltage is at least 80% of the source voltage.

Ans. 9.9 mV; 9.1 mV; 5 mV; 0.9 mV; 4 kΩ

1.4 A signal source that is most conveniently represented by its Norton equivalent form has $i_s = 10$ μA and $R_s = 100$ kΩ. If the source feeds a load resistance R_L, find the current i_o that flows through the load for $R_L = 1$ kΩ, 10 kΩ, 100 kΩ, and 1 MΩ. Also, find the largest permissible value of R_L for which the load current is at least 80% of the source current.

Ans. 9.9 μA; 9.1 μA; 5 μA; 0.9 μA; 25 kΩ

From the discussion above, it should be apparent that a signal is a time-varying quantity that can be represented by a graph such as that shown in Fig. 1.3. In fact, the information content of the signal is represented by the changes in its magnitude as time progresses; that is, the information is contained in the "wiggles" in the signal waveform. In general, such waveforms are difficult to characterize mathematically. In other words, it is not easy to describe succinctly an arbitrary-looking waveform such as that of Fig. 1.3. Of course, such a description is of great importance for the purpose of designing appropriate signal-processing circuits that perform desired functions on the given signal. An effective approach to signal characterization is studied in the next section.

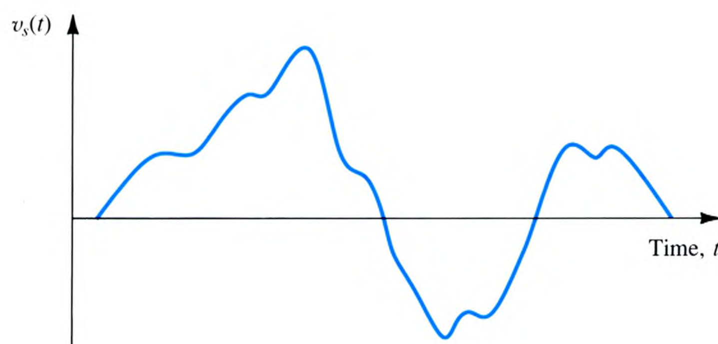


Figure 1.3 An arbitrary voltage signal $v_s(t)$.

1.2 Frequency Spectrum of Signals

It can be extremely useful to characterize a signal, and for that matter any arbitrary function of time, in terms of its **frequency spectrum**. We can obtain such a description of signals through the mathematical tools of **Fourier series** and **Fourier transform**.[1] We are not interested here in the details of these transformations; suffice it to say that they provide the means for representing a voltage signal $v_s(t)$ or a current signal $i_s(t)$ as the sum of sine-wave signals of different frequencies and amplitudes. This makes the sine wave a very important signal in the analysis, design, and testing of electronic circuits. Therefore, we shall briefly review the properties of the sinusoid.

Figure 1.4 shows a sine-wave voltage signal $v_a(t)$,

$$v_a(t) = V_a \sin\omega t \tag{1.1}$$

where V_a denotes the peak value or amplitude in volts and ω denotes the angular frequency in radians per second; that is, $\omega = 2\pi f$ rad/s, where f is the frequency in hertz, $f = 1/T$ Hz, and T is the period in seconds.

The sine-wave signal is completely characterized by its peak value V_a, its frequency ω, and its phase with respect to an arbitrary reference time. In the case depicted in Fig. 1.4, the time origin has been chosen so that the phase angle is 0. It is common to express the amplitude of a sine-wave signal in terms of its root-mean-square (rms) value, which is equal to the peak value divided by $\sqrt{2}$. Thus the rms value of the sinusoid $v_a(t)$ of Fig. 1.4 is $V_a/\sqrt{2}$. For instance, when we speak of the wall power supply in our homes as being 120 V, we mean that it has a sine waveform of $120\sqrt{2}$ volts peak value.

Returning now to the representation of signals as the sum of sinusoids, we note that the Fourier series is utilized to accomplish this task for the special case of a signal that is a periodic function of time. On the other hand, the Fourier transform is more general and can be used to obtain the frequency spectrum of a signal whose waveform is an arbitrary function of time.

The Fourier series allows us to express a given periodic function of time as the sum of an infinite number of sinusoids whose frequencies are harmonically related. For instance, the symmetrical square-wave signal in Fig. 1.5 can be expressed as

$$v(t) = \frac{4V}{\pi}\left(\sin\omega_0 t + \frac{1}{3}\sin 3\omega_0 t + \frac{1}{5}\sin 5\omega_0 t + \cdots\right) \tag{1.2}$$

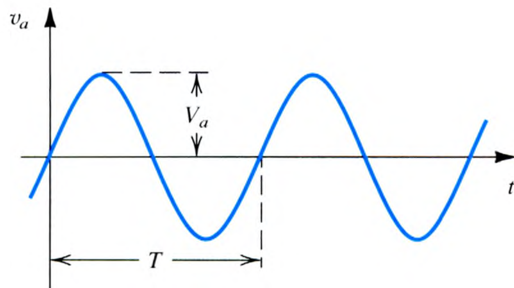


Figure 1.4 Sine-wave voltage signal of amplitude V_a and frequency $f = 1/T$ Hz. The angular frequency $\omega = 2\pi f$ rad/s.

[1]The reader who has not yet studied these topics should not be alarmed. No detailed application of this material will be made until Chapter 10. Nevertheless, a general understanding of Section 1.2 should be very helpful in studying early parts of this book.

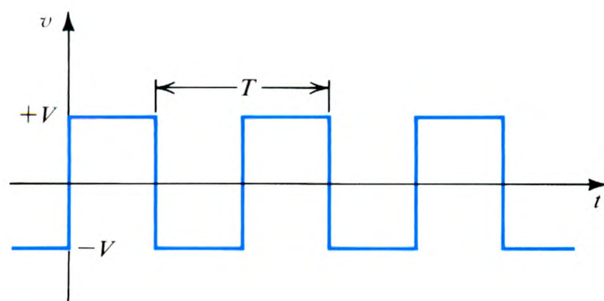


Figure 1.5 A symmetrical square-wave signal of amplitude V.

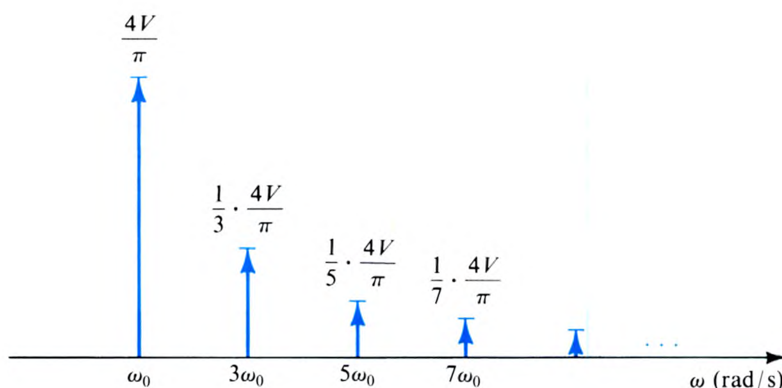


Figure 1.6 The frequency spectrum (also known as the **line spectrum**) of the periodic square wave of Fig. 1.5.

where V is the amplitude of the square wave and $\omega_0 = 2\pi/T$ (T is the period of the square wave) is called the **fundamental frequency**. Note that because the amplitudes of the harmonics progressively decrease, the infinite series can be truncated, with the truncated series providing an approximation to the square waveform.

The sinusoidal components in the series of Eq. (1.2) constitute the frequency spectrum of the square-wave signal. Such a spectrum can be graphically represented as in Fig. 1.6, where the horizontal axis represents the angular frequency ω in radians per second.

The Fourier transform can be applied to a nonperiodic function of time, such as that depicted in Fig. 1.3, and provides its frequency spectrum as a continuous function of frequency, as indicated in Fig. 1.7. Unlike the case of periodic signals, where the spectrum consists of discrete frequencies (at ω_0 and its harmonics), the spectrum of a nonperiodic signal contains in general all possible frequencies. Nevertheless, the essential parts of the spectra of practical signals are usually confined to relatively short segments of the frequency (ω) axis—an observation that is very useful in the processing of such signals. For instance, the spectrum of audible sounds such as speech and music extends from about 20 Hz to about 20 kHz—a frequency range known as the **audio band**. Note that although some musical tones have frequencies above 20 kHz, the human ear is incapable of hearing frequencies that are much above 20 kHz. Analog video signals have their spectra in the range of 0 MHz to 4.5 MHz.

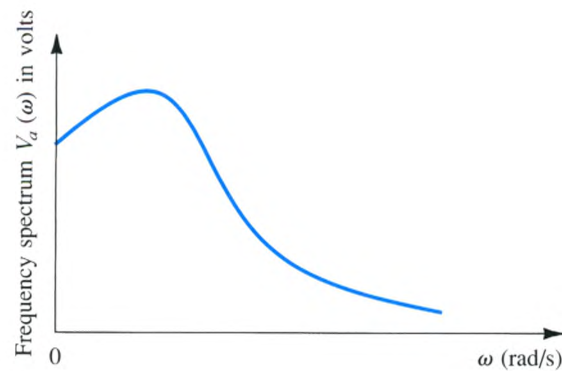


Figure 1.7 The frequency spectrum of an arbitrary waveform such as that in Fig. 1.3.

We conclude this section by noting that a signal can be represented either by the manner in which its waveform varies with time, as for the voltage signal $v_a(t)$ shown in Fig. 1.3, or in terms of its frequency spectrum, as in Fig. 1.7. The two alternative representations are known as the time-domain representation and the frequency-domain representation, respectively. The frequency-domain representation of $v_a(t)$ will be denoted by the symbol $V_a(\omega)$.

EXERCISES

1.5 Find the frequencies f and ω of a sine-wave signal with a period of 1 ms.
Ans. $f = 1000$ Hz; $\omega = 2\pi \times 10^3$ rad/s

1.6 What is the period T of sine waveforms characterized by frequencies of (a) $f = 60$ Hz? (b) $f = 10^{-3}$ Hz? (c) $f = 1$ MHz?
Ans. 16.7 ms; 1000 s; 1 μs

1.7 The UHF (ultra high frequency) television broadcast band begins with channel 14 and extends from 470 MHz to 608 MHz. If 6 MHz is allocated for each channel, how many channels can this band accommodate?
Ans. 23; channels 14 to 36

1.8 When the square-wave signal of Fig. 1.5, whose Fourier series is given in Eq. (1.2), is applied to a resistor, the total power dissipated may be calculated directly using the relationship $P = 1/T \int_0^T (v^2/R)\,dt$ or indirectly by summing the contribution of each of the harmonic components, that is, $P = P_1 + P_3 + P_5 + \ldots$, which may be found directly from rms values. Verify that the two approaches are equivalent. What fraction of the energy of a square wave is in its fundamental? In its first five harmonics? In its first seven? First nine? In what number of harmonics is 90% of the energy? (Note that in counting harmonics, the fundamental at ω_0 is the first, the one at $2\omega_0$ is the second, etc.)
Ans. 0.81; 0.93; 0.95; 0.96; 3

1.3 Analog and Digital Signals

The voltage signal depicted in Fig. 1.3 is called an **analog signal**. The name derives from the fact that such a signal is *analogous* to the physical signal that it represents. The magnitude of an analog signal can take on any value; that is, the amplitude of an analog signal exhibits a continuous variation over its range of activity. The vast majority of signals in the world around us are analog. Electronic circuits that process such signals are known as **analog circuits**. A variety of analog circuits will be studied in this book.

An alternative form of signal representation is that of a sequence of numbers, each number representing the signal magnitude at an instant of time. The resulting signal is called a **digital signal**. To see how a signal can be represented in this form—that is, how signals can be converted from analog to digital form—consider Fig. 1.8(a). Here the curve represents a voltage signal, identical to that in Fig. 1.3. At equal intervals along the time axis, we have marked the time instants t_0, t_1, t_2, and so on. At each of these time instants, the magnitude of the signal is measured, a process known as **sampling**. Figure 1.8(b) shows a representation of the signal of Fig. 1.8(a) in terms of its samples. The signal of Fig. 1.8(b) is defined only at the sampling instants; it no longer is a continuous function of time; rather, it is a **discrete-time signal**. However, since the magnitude of each sample can take any value in a continuous range, the signal in Fig. 1.8(b) is still an analog signal.

Now if we represent the magnitude of each of the signal samples in Fig. 1.8(b) by a number having a finite number of digits, then the signal amplitude will no longer be continuous; rather, it is said to be **quantized**, **discretized**, or **digitized**. The resulting digital signal then is simply a sequence of numbers that represent the magnitudes of the successive signal samples.

The choice of number system to represent the signal samples affects the type of digital signal produced and has a profound effect on the complexity of the digital circuits required to process the signals. It turns out that the **binary** number system results in the simplest possible digital signals and circuits. In a binary system, each digit in the number takes on one of only two possible values, denoted 0 and 1. Correspondingly, the digital signals in binary systems need have only two voltage levels, which can be labeled low and high. As an example, in some of the digital circuits studied in this book, the levels may be 0 V and +1.8 V. Figure 1.9 shows the time variation of such a digital signal. Observe that the waveform is a pulse train with 0 V representing a 0 signal, or logic 0, and +1.8 V representing logic 1. Unlike the original analog signal, which can take on any real value and therefore can be corrupted by noise, the digital waveform can withstand some noise while still being able to distinguish between logic levels without any loss of information.

If we use N *b*inary dig*its* (bits) to represent each sample of the analog signal, then the digitized sample value can be expressed as

$$D = b_0 2^0 + b_1 2^1 + b_2 2^2 + \cdots + b_{N-1} 2^{N-1} \tag{1.3}$$

where $b_0, b_1, \ldots, b_{N-1}$, denote the N bits and have values of 0 or 1. Here bit b_0 is the **least significant bit** (**LSB**), and bit b_{N-1} is the **most significant bit** (**MSB**). Conventionally, this binary number is written as $b_{N-1}\, b_{N-2} \ldots b_0$. We observe that such a representation quantizes the analog sample into one of 2^N levels. Obviously the greater the number of bits (i.e., the larger the N), the closer the digital word D approximates the magnitude of the analog sample. That is, increasing the number of bits reduces the *quantization error* and increases the resolution of the analog-to-digital conversion. This improvement is, however, usually obtained at the expense of more complex and hence more costly circuit implementations. It

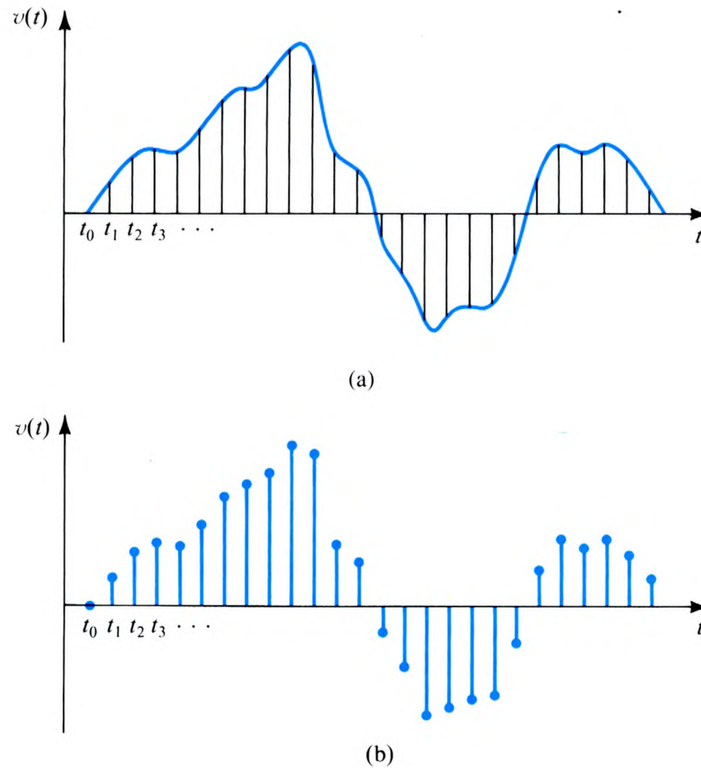


Figure 1.8 Sampling the continuous-time analog signal in **(a)** results in the discrete-time signal in **(b)**.

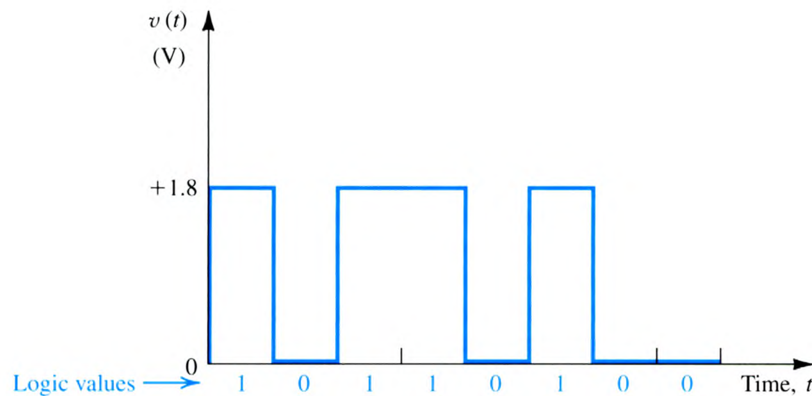


Figure 1.9 Variation of a particular binary digital signal with time.

is not our purpose here to delve into this topic any deeper; we merely want the reader to appreciate the nature of analog and digital signals. Nevertheless, it is an opportune time to introduce a very important circuit building block of modern electronic systems: the **analog-to-digital converter** (**A/D** or **ADC**) shown in block form in Fig. 1.10. The ADC accepts at its input the samples of an analog signal and provides for each input sample the corresponding N-bit digital representation (according to Eq. 1.3) at its N output terminals.

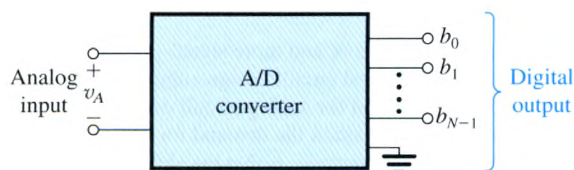


Figure 1.10 Block-diagram representation of the analog-to-digital converter (ADC).

Thus although the voltage at the input might be, say, 1.51 V, at each of the output terminals (say, at the ith terminal), the voltage will be either low (0 V) or high (1.8 V) if b_i is supposed to be 0 or 1, respectively. The dual circuit of the ADC is the **digital-to-analog converter** (**D/A** or **DAC**). It converts an N-bit digital input to an analog output voltage.

Once the signal is in digital form, it can be processed using **digital circuits**. Of course digital circuits can deal also with signals that do not have an analog origin, such as the signals that represent the various instructions of a digital computer.

Since digital circuits deal exclusively with binary signals, their design is simpler than that of analog circuits. Furthermore, digital systems can be designed using a relatively few different kinds of digital circuit blocks. However, a large number (e.g., hundreds of thousands or even millions) of each of these blocks are usually needed. Thus the design of digital circuits poses its own set of challenges to the designer but provides reliable and economic implementations of a great variety of signal-processing functions, many of which are not possible with analog circuits. Many signal-processing functions that relied upon analog circuits in the past are now being performed digitally. Examples around us abound, from the digital watch and calculator to digital audio systems and telephony. Modern computers and smartphones are enabled by very-large-scale digital circuits. Image and video recording, storage, and transmission are all predominantly performed by digital circuits. Digital circuits have a particularly special role to play in communication because digital information is inherently more robust to noise than an analog signal.

The basic building blocks of digital systems are logic circuits and memory circuits. We will study both in this book, beginning in Chapter 16.

One final remark: Although the digital processing of signals may appear to be all-pervasive, in fact many electronic systems include both analog and digital parts. It follows that a good electronics engineer must be proficient in the design of both analog and digital circuits, or **mixed-signal** or **mixed-mode** design as it is currently known. Such is the aim of this book.

EXERCISE

1.9 Consider a 4-bit digital word $D = b_3b_2b_1b_0$ (see Eq. 1.3) used to represent an analog signal v_A that varies between 0 V and +3.75 V.

(a) Give D corresponding to $v_A = 0$ V, 0.25 V, 1 V, and 3.75 V.

(b) What change in v_A causes a change from 0 to 1 in (i) b_0, (ii) b_1, (iii) b_2, and (iv) b_3?

(c) If $v_A = 1.3$ V, what do you expect D to be? What is the resulting error in representation?

Ans. (a) 0000, 0001, 0100, 1111; (b) +0.25 V, +0.50 V, +1 V, +2 V; (c) 0101, –4%

ANALOG VS. DIGITAL CIRCUIT ENGINEERS

As digital became the preferred implementation of more and more signal-processing functions, the need arose for greater numbers of digital circuit design engineers. Yet despite predictions made periodically that the demand for analog circuit design engineers would lessen, this has not been the case. Rather, the demand for analog engineers has, if anything, increased. What is true, however, is that the skill level required of analog engineers has risen. Not only are they asked to design circuits of greater sophistication and tighter specifications, but they also have to do this using technologies that are optimized for digital (and not analog) circuits. This is dictated by economics, as digital usually constitutes the larger part of most systems.

1.4 Amplifiers

In this section, we shall introduce the most fundamental signal-processing function, one that is employed in some form in almost every electronic system, namely, signal amplification. We shall study the amplifier as a circuit building block; that is, we shall consider its external characteristics and leave the design of its internal circuit to later chapters.

1.4.1 Signal Amplification

From a conceptual standpoint, the simplest signal-processing task is **signal amplification**. The need for amplification arises because transducers provide signals that are said to be "weak," that is, in the microvolt (μV) or millivolt (mV) range and possessing little energy. Such signals are too small for reliable processing, which becomes much easier if the signal magnitude is made larger. The functional block that accomplishes this task is the **signal amplifier**.

It is appropriate at this point to discuss the need for **linearity** in amplifiers. Care must be exercised in the amplification of a signal, so that the information contained in the signal is not changed and no new information is introduced. Thus when we feed the signal shown in Fig. 1.3 to an amplifier, we want the output signal of the amplifier to be an exact replica of that at the input, except of course for having larger magnitude. In other words, the "wiggles" in the output waveform must be identical to those in the input waveform. Any change in waveform is considered to be **distortion** and is obviously undesirable.

An amplifier that preserves the details of the signal waveform is characterized by the relationship

$$v_o(t) = A v_i(t) \tag{1.4}$$

where v_i and v_o are the input and output signals, respectively, and A is a constant representing the magnitude of amplification, known as **amplifier gain**. Equation (1.4) is a linear relationship; hence the amplifier it describes is a **linear amplifier**. It should be easy to see that if the relationship between v_o and v_i contains higher powers of v_i, then the waveform of v_o will no longer be identical to that of v_i. The amplifier is then said to exhibit **nonlinear distortion**.

The amplifiers discussed so far are primarily intended to operate on very small input voltage signals. Their purpose is to make the signal magnitude larger, and therefore they are

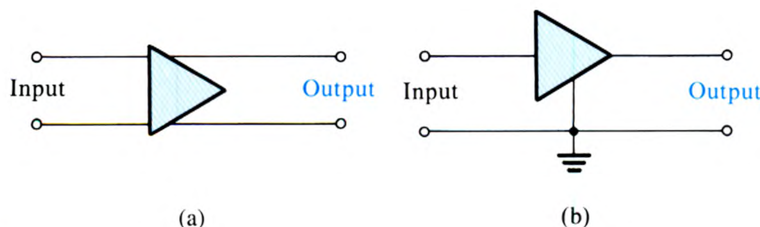


Figure 1.11 **(a)** Circuit symbol for amplifier. **(b)** An amplifier with a common terminal (ground) between the input and output ports.

thought of as **voltage amplifiers**. The **preamplifier** in the home stereo system is an example of a voltage amplifier.

At this time we wish to mention another type of amplifier, namely, the **power amplifier**. Such an amplifier may provide only a modest amount of voltage gain but substantial current gain. Thus while absorbing little power from the input signal source to which it is connected, often a preamplifier, it delivers large amounts of power to its load. An example is found in the power amplifier of the home stereo system, whose purpose is to provide sufficient power to drive the loudspeaker, which is the amplifier load. Here we should note that the loudspeaker is the output transducer of the stereo system; it converts the electric output signal of the system into an acoustic signal. A further appreciation of the need for linearity can be acquired by reflecting on the power amplifier. A linear power amplifier causes both soft and loud music passages to be reproduced without distortion.

1.4.2 Amplifier Circuit Symbol

The signal amplifier is obviously a two-port circuit. Its function is conveniently represented by the circuit symbol of Fig. 1.11(a). This symbol clearly distinguishes the input and output ports and indicates the direction of signal flow. Thus, in subsequent diagrams it will not be necessary to label the two ports "input" and "output." For generality we have shown the amplifier to have two input terminals that are distinct from the two output terminals. A more common situation is illustrated in Fig. 1.11(b), where a common terminal exists between the input and output ports of the amplifier. This common terminal is used as a reference point and is called the **circuit ground**.

1.4.3 Voltage Gain

A linear amplifier accepts an input signal $v_I(t)$ and provides at the output, across a load resistance R_L (see Fig. 1.12(a)), an output signal $v_O(t)$ that is a magnified replica of $v_I(t)$. The **voltage gain** of the amplifier is defined by

$$\text{Voltage gain } (A_v) = \frac{v_O}{v_I} \tag{1.5}$$

Fig. 1.12(b) shows the **transfer characteristic** of a linear amplifier. If we apply to the input of this amplifier a sinusoidal voltage of amplitude $\hat{V}$, we obtain at the output a sinusoid of amplitude $A_v\hat{V}$.

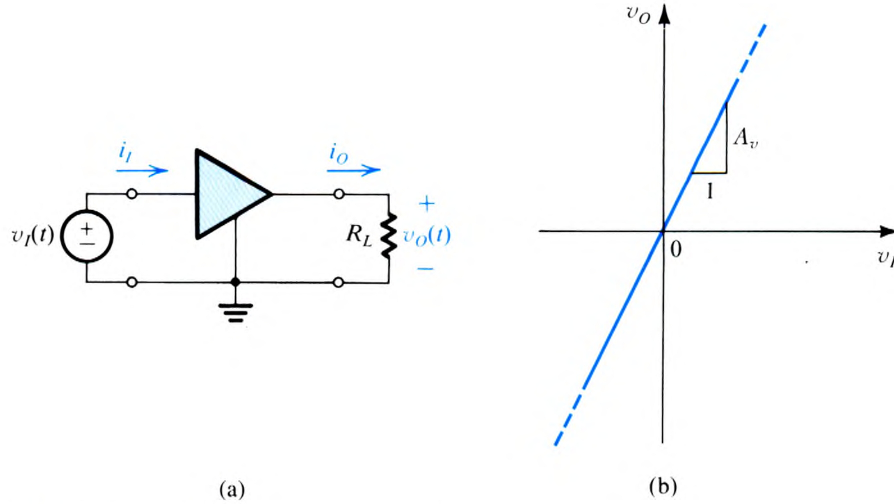


Figure 1.12 **(a)** A voltage amplifier fed with a signal $v_I(t)$ and connected to a load resistance R_L. **(b)** Transfer characteristic of a linear voltage amplifier with voltage gain A_v.

1.4.4 Power Gain and Current Gain

An amplifier increases the signal power, an important feature that distinguishes an amplifier from a transformer. In the case of a transformer, although the voltage delivered to the load could be greater than the voltage feeding the input side (the primary), the power delivered to the load (from the secondary side of the transformer) is less than or at most equal to the power supplied by the signal source. On the other hand, an amplifier provides the load with power greater than that obtained from the signal source. That is, amplifiers have power gain. The **power gain** of the amplifier in Fig. 1.12(a) is defined as

$$\text{Power gain } (A_p) \equiv \frac{\text{load power } (P_L)}{\text{input power}(P_I)} \tag{1.6}$$

$$= \frac{v_O i_O}{v_I i_I} \tag{1.7}$$

where i_O is the current that the amplifier delivers to the load $(R_L), i_O = v_O/R_L$, and i_I is the current the amplifier draws from the signal source. The **current gain** of the amplifier is defined as

$$\text{Current gain } (A_i) \equiv \frac{i_O}{i_I} \tag{1.8}$$

From Eqs. (1.5) to (1.8) we note that

$$A_p = A_v A_i \tag{1.9}$$

1.4.5 Expressing Gain in Decibels

The amplifier gains defined above are ratios of similarly dimensioned quantities. Thus they will be expressed either as dimensionless numbers or, for emphasis, as V/V for the voltage gain, A/A for the current gain, and W/W for the power gain. Alternatively, for a

number of reasons, some of them historic, electronics engineers express amplifier gain with a logarithmic measure. Specifically the voltage gain A_v can be expressed as

$$\text{Voltage gain in decibels} = 20 \log|A_v| \quad \text{dB}$$

and the current gain A_i can be expressed as

$$\text{Current gain in decibels} = 20 \log|A_i| \quad \text{dB}$$

Since power is related to voltage (or current) squared, the power gain A_p can be expressed in decibels as

$$\text{Power gain in decibels} = 10 \log A_p \quad \text{dB}$$

The absolute values of the voltage and current gains are used because in some cases A_v or A_i will be a negative number. A negative gain A_v simply means that there is a 180° phase difference between input and output signals; it does not imply that the amplifier is **attenuating** the signal. On the other hand, an amplifier whose voltage gain is, say, −20 dB is in fact attenuating the input signal by a factor of 10 (i.e., $A_v = 0.1$ V/V).

1.4.6 The Amplifier Power Supplies

Since the power delivered to the load is greater than the power drawn from the signal source, you may wonder where this additional power comes from. The answer is found by observing that amplifiers need dc power supplies for their operation. These dc sources supply the extra power delivered to the load as well as any power that might be dissipated in the internal circuit of the amplifier (such power is converted to heat). In Fig. 1.12(a) we have not explicitly shown these dc sources.

Figure 1.13(a) shows an amplifier that requires two dc sources: one positive of value V_{CC} and one negative of value V_{EE}. The amplifier has two terminals, labeled V^+ and V^-, for connection to the dc supplies. For the amplifier to operate, the terminal labeled V^+ has to be connected to the positive side of a dc source whose voltage is V_{CC} and whose negative side is connected to the circuit ground. Also, the terminal labeled V^- has to be connected to the

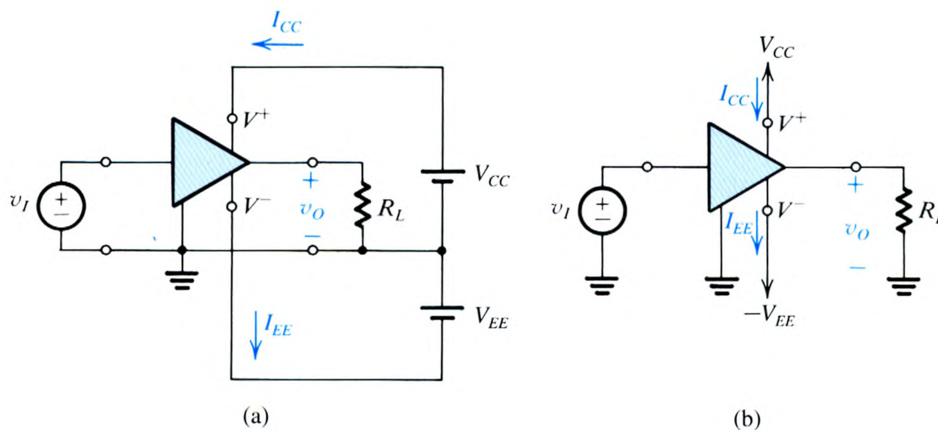


Figure 1.13 An amplifier that requires two dc supplies (shown as batteries) for operation.

negative side of a dc source whose voltage is V_{EE} and whose positive side is connected to the circuit ground. Now, if the current drawn from the positive supply is denoted I_{CC} and that from the negative supply is I_{EE} (see Fig. 1.13a), then the dc power delivered to the amplifier is

$$P_{dc} = V_{CC}I_{CC} + V_{EE}I_{EE}$$

If the power dissipated in the amplifier circuit is denoted $P_{dissipated}$, the power-balance equation for the amplifier can be written as

$$P_{dc} + P_I = P_L + P_{dissipated}$$

where P_I is the power drawn from the signal source and P_L is the power delivered to the load. Since the power drawn from the signal source is usually small, the amplifier power **efficiency** is defined as

$$\eta \equiv \frac{P_L}{P_{dc}} \times 100 \tag{1.10}$$

The power efficiency is an important performance parameter for amplifiers that handle large amounts of power. Such amplifiers, called power amplifiers, are used, for example, as output amplifiers of stereo systems.

In order to simplify circuit diagrams, we shall adopt the convention illustrated in Fig. 1.13(b). Here the V^+ terminal is shown connected to an arrowhead pointing upward and the V^- terminal to an arrowhead pointing downward. The corresponding voltage is indicated next to each arrowhead. Note that in many cases we will not explicitly show the connections of the amplifier to the dc power sources. Finally, we note that some amplifiers require only one power supply.

Example 1.2

Consider a microphone producing a sinusoidal signal that is 400-mV peak. It delivers 10-μA peak sinusoidal current to an amplifier that operates from ±1-V power supplies. The amplifier delivers a 0.8-V peak sinusoid to a speaker load with 32-Ω resistance. The amplifier draws a current of 30 mA from each of its two power supplies. Find the voltage gain, the current gain, the power gain, the power drawn from the dc supplies, the power dissipated in the amplifier, and the amplifier efficiency.

Solution

$$A_v = \frac{0.8 \text{ V}}{0.4 \text{ V}} = 2 \text{ V/V, or } A_v = 20 \log 2 = 6 \text{ dB}$$

$$\hat{I}_o = \frac{0.8 \text{ V}}{32 \ \Omega} = 25 \text{ mA}$$

$$A_i = \frac{\hat{I}_o}{\hat{I}_i} = \frac{25 \text{ mA}}{0.01 \text{ mA}} = 2500 \text{ A/A, or } A_i = 20 \log 2500 = 68 \text{ dB}$$

$$P_L = V_{o_{\text{rms}}} I_{o_{\text{rms}}} = \frac{0.8 \text{ V}}{\sqrt{2}} \frac{25 \text{ mA}}{\sqrt{2}} = 10 \text{ mW}$$

$$P_I = V_{i_{\text{rms}}} I_{i_{\text{rms}}} = \frac{0.4 \text{ V}}{\sqrt{2}} \frac{0.01 \text{ mA}}{\sqrt{2}} = 2 \ \mu\text{W}$$

$$A_p = \frac{P_L}{P_I} = \frac{10 \text{ mW}}{2 \ \mu\text{W}} = 5000 \text{ W/W, or } A_p = 10 \log 5000 = 37 \text{ dB}$$

$$P_{\text{dc}} = 1 \text{ V} \times 30 \text{ mA} + 1 \text{ V} \times 30 \text{ mA} = 60 \text{ mW}$$

$$P_{\text{dissipated}} = P_{\text{dc}} + P_I - P_L$$

$$= 60 \text{ mW} + 0.002 \text{ mW} - 10 \text{ mW} \simeq 50 \text{ mW}$$

$$\eta = \frac{P_L}{P_{\text{dc}}} \times 100 = 16.7\%$$

From the above example we observe that the amplifier converts some of the dc power it draws from the power supplies to signal power that it delivers to the load.

1.4.7 Amplifier Saturation

Practically speaking, the amplifier transfer characteristic remains linear over only a limited range of input and output voltages. For an amplifier operated from two power supplies the output voltage cannot exceed a specified positive limit and cannot decrease below a specified negative limit. The resulting transfer characteristic is shown in Fig. 1.14, with the positive and negative saturation levels denoted L_+ and L_-, respectively. Each of the two saturation levels is usually within a fraction of a volt of the voltage of the corresponding power supply.

Obviously, in order to avoid distorting the output signal waveform, the input signal swing must be kept within the linear range of operation,

$$\frac{L_-}{A_v} \le v_I \le \frac{L_+}{A_v}$$

In Fig. 1.14, which shows two input waveforms and the corresponding output waveforms, the peaks of the larger waveform have been clipped off because of amplifier saturation.

1.4.8 Symbol Convention

At this point, we draw your attention to the terminology we will use throughout the book. To illustrate, Fig. 1.15 shows the waveform of a current $i_C(t)$ that is flowing through a branch in a particular circuit. The current $i_C(t)$ consists of a dc component I_C on which is superimposed a sinusoidal component $i_c(t)$ whose peak amplitude is I_c. Observe that at a time t, the **total**

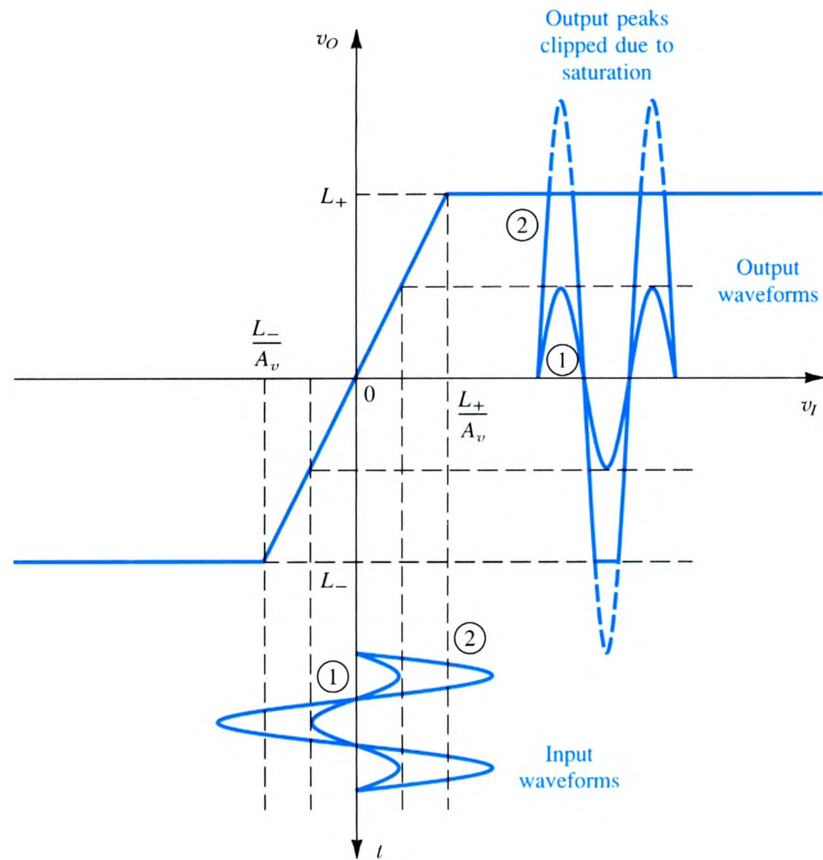


Figure 1.14 An amplifier transfer characteristic that is linear except for output saturation.

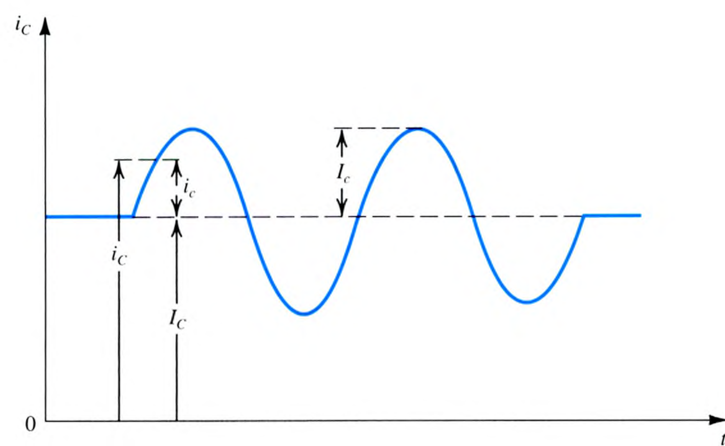


Figure 1.15 Symbol convention employed throughout the book.

instantaneous current $i_C(t)$ is the sum of the dc current I_C and the signal current $i_c(t)$,

$$i_C(t) = I_C + i_c(t) \tag{1.11}$$

where the signal current is given by

$$i_c(t) = I_c \sin \omega t$$

Thus, we state some conventions: Total instantaneous quantities are denoted by a lowercase symbol with uppercase subscript(s), for example, $i_C(t), v_{DS}(t)$. Direct-current (dc) quantities are denoted by an uppercase symbol with uppercase subscript(s), for example, I_C, V_{DS}. Incremental signal quantities are denoted by a lowercase symbol with lowercase subscript(s), for example, $i_c(t), v_{gs}(t)$. If the signal is a sine wave, then its amplitude is denoted by an uppercase symbol with lowercase subscript(s), for example, I_c, V_{gs}. Finally, although not shown in Fig. 1.15, dc power supplies are denoted by an uppercase letter with a double-letter uppercase subscript, for example, V_{CC}, V_{DD}. A similar notation is used for the dc current drawn from the power supply, for example, I_{CC}, I_{DD}.

EXERCISES

1.10 An amplifier has a voltage gain of 100 V/V and a current gain of 1000 A/A. Express the voltage and current gains in decibels and find the power gain.
Ans. 40 dB; 60 dB; 50 dB

1.11 An amplifier operating from a single 15-V supply provides a 12-V peak-to-peak sine-wave signal to a 1-kΩ load and draws negligible input current from the signal source. The dc current drawn from the 15-V supply is 8 mA. What is the power dissipated in the amplifier, and what is the amplifier efficiency?
Ans. 120 mW; 15%

1.5 Circuit Models for Amplifiers

A substantial part of this book is concerned with the design of amplifier circuits that use transistors of various types. Such circuits will vary in complexity from those using a single transistor to those with 20 or more devices. In order to be able to apply the resulting amplifier circuit as a building block in a system, one must be able to characterize, or **model**, its terminal behavior. In this section, we study simple but effective amplifier models. These models apply irrespective of the complexity of the internal circuit of the amplifier. The values of the model parameters can be found either by analyzing the amplifier circuit or by performing measurements at the amplifier terminals.

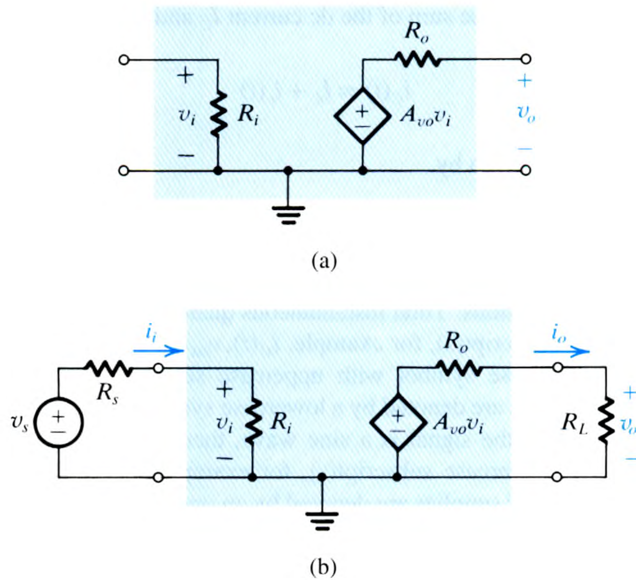


Figure 1.16 **(a)** Circuit model for the voltage amplifier. **(b)** The voltage amplifier with input signal source and load.

1.5.1 Voltage Amplifiers

Figure 1.16(a) shows a circuit model for the voltage amplifier. The model consists of a voltage-controlled voltage source having a gain factor A_{vo}, an input resistance R_i that accounts for the fact that the amplifier draws an input current from the signal source, and an output resistance R_o that accounts for the change in output voltage as the amplifier is called upon to supply output current to a load. To be specific, we show in Fig. 1.16(b) the amplifier model fed with a signal voltage source v_s having a resistance R_s and connected at the output to a load resistance R_L. The nonzero output resistance R_o causes only a fraction of $A_{vo}v_i$ to appear across the output. Using the voltage-divider rule we obtain

$$v_o = A_{vo}v_i \frac{R_L}{R_L + R_o}$$

Thus the voltage gain is given by

$$A_v \equiv \frac{v_o}{v_i} = A_{vo} \frac{R_L}{R_L + R_o} \tag{1.12}$$

It follows that in order not to lose gain in coupling the amplifier output to a load, the output resistance R_o should be much smaller than the load resistance R_L. In other words, for a given R_L one must design the amplifier so that its R_o is much smaller than R_L. Furthermore, there are applications in which R_L is known to vary over a certain range. In order to keep the output voltage v_o as constant as possible, the amplifier is designed with R_o much smaller than the lowest value of R_L. An ideal voltage amplifier is one with $R_o = 0$. Equation (1.12) indicates also that for $R_L = \infty, A_v = A_{vo}$. Thus A_{vo} is the voltage gain of the unloaded amplifier, or the **open-circuit voltage gain**. It should also be clear that in specifying the voltage gain of an amplifier, one must also specify the value of load resistance at which this gain is measured or

calculated. If a load resistance is not specified, it is normally assumed that the given voltage gain is the open-circuit gain A_{vo}.

The finite input resistance R_i introduces another voltage-divider action at the input, with the result that only a fraction of the source signal v_s actually reaches the input terminals of the amplifier; that is,

$$v_i = v_s \frac{R_i}{R_i + R_s} \tag{1.13}$$

It follows that in order not to lose a significant portion of the input signal in coupling the signal source to the amplifier input, the amplifier must be designed to have an input resistance R_i much greater than the resistance of the signal source, $R_i \gg R_s$. Furthermore, there are applications in which the source resistance is known to vary over a certain range. To minimize the effect of this variation on the value of the signal that appears at the input of the amplifier, the designer ensures that R_i is much greater than the largest value of R_s. An ideal voltage amplifier is one with $R_i = \infty$. In this ideal case both the current gain and power gain become infinite.

The overall voltage gain (v_o/v_s) can be found by combining Eqs. (1.12) and (1.13),

$$\frac{v_o}{v_s} = A_{vo} \frac{R_i}{R_i + R_s} \frac{R_L}{R_L + R_o}$$

There are situations in which one is interested not in voltage gain but only in a significant power gain. For instance, the source signal can have a respectable voltage but a source resistance that is much greater than the load resistance. Connecting the source directly to the load would result in significant signal attenuation. In such a case, one requires an amplifier with a high input resistance (much greater than the source resistance) and a low output resistance (much smaller than the load resistance) but with a modest voltage gain (or even unity gain). Such an amplifier is referred to as a **buffer amplifier**. We shall encounter buffer amplifiers often throughout this book.

EXERCISES

1.12 A sensor producing a voltage of 1 V rms with a source resistance of 1 MΩ is available to drive a 10-Ω load. If connected directly, what voltage and power levels result at the load? If a unity-gain (i.e., A_{vo} = 1) buffer amplifier with 1-MΩ input resistance and 10-Ω output resistance is interposed between source and load, what do the output voltage and power levels become? For the new arrangement, find the voltage gain from source to load, and the power gain (both expressed in decibels).
Ans. 10 μV rms; 10^{-11} W; 0.25 V rms; 6.25 mW; −12 dB; 44 dB

1.13 The output voltage of a voltage amplifier has been found to decrease by 20% when a load resistance of 1 kΩ is connected. What is the value of the amplifier output resistance?
Ans. 250 Ω

1.14 An amplifier with an open-circuit voltage gain of +40 dB, an input resistance of 10 kΩ, and an output resistance of 1 kΩ is used to drive a 1-kΩ load. What is the value of A_{vo}? Find the value of the power gain in decibels.
Ans. 100 V/V; 44 dB

1.5.2 Cascaded Amplifiers

To meet given amplifier specifications, we often need to design the amplifier as a cascade of two or more stages. The stages are usually not identical; rather, each is designed to serve a specific purpose. For instance, in order to provide the overall amplifier with a large input resistance, the first stage is usually required to have a large input resistance. Also, in order to equip the overall amplifier with a low output resistance, the final stage in the cascade is usually designed to have a low output resistance. To illustrate the analysis and design of cascaded amplifiers, we consider a practical example.

Example 1.3

Figure 1.17 depicts an amplifier composed of a cascade of three stages. The amplifier is fed by a signal source with a source resistance of 100 kΩ and delivers its output into a load resistance of 100 Ω. The first stage has a relatively high input resistance and a modest gain factor of 10. The second stage has a higher gain factor but lower input resistance. Finally, the last, or output, stage has unity gain but a low output resistance. We wish to evaluate the overall voltage gain, that is, v_L/v_s, the current gain, and the power gain.

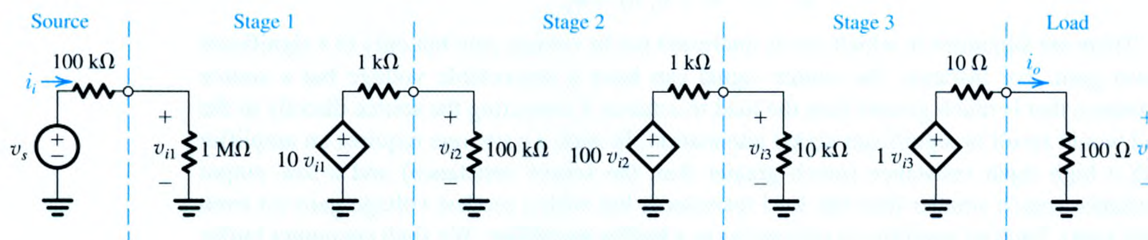


Figure 1.17 Three-stage amplifier for Example 1.3.

Solution

The fraction of source signal appearing at the input terminals of the amplifier is obtained using the voltage-divider rule at the input, as follows:

$$\frac{v_{i1}}{v_s} = \frac{1\text{ M}\Omega}{1\text{ M}\Omega + 100\text{ k}\Omega} = 0.909\text{ V/V}$$

The voltage gain of the first stage is obtained by considering the input resistance of the second stage to be the load of the first stage; that is,

$$A_{v1} \equiv \frac{v_{i2}}{v_{i1}} = 10\frac{100\text{ k}\Omega}{100\text{ k}\Omega + 1\text{ k}\Omega} = 9.9\text{ V/V}$$

Similarly, the voltage gain of the second stage is obtained by considering the input resistance of the third stage to be the load of the second stage,

$$A_{v2} \equiv \frac{v_{i3}}{v_{i2}} = 100\frac{10\text{ k}\Omega}{10\text{ k}\Omega + 1\text{ k}\Omega} = 90.9\text{ V/V}$$

Finally, the voltage gain of the output stage is as follows:

$$A_{v3} \equiv \frac{v_L}{v_{i3}} = 1 \frac{100\ \Omega}{100\ \Omega + 10\ \Omega} = 0.909 \text{ V/V}$$

The total gain of the three stages in cascade can now be found from

$$A_v \equiv \frac{v_L}{v_{i1}} = A_{v1}A_{v2}A_{v3} = 818 \text{ V/V}$$

or 58.3 dB.

To find the voltage gain from source to load, we multiply A_v by the factor representing the loss of gain at the input; that is,

$$\frac{v_L}{v_s} = \frac{v_L}{v_{i1}}\frac{v_{i1}}{v_s} = A_v \frac{v_{i1}}{v_s}$$

$$= 818 \times 0.909 = 743.6 \text{ V/V}$$

or 57.4 dB.

The current gain is found as follows:

$$A_i \equiv \frac{i_o}{i_i} = \frac{v_L/100\ \Omega}{v_{i1}/1\ \text{M}\Omega}$$

$$= 10^4 \times A_v = 8.18 \times 10^6 \text{ A/A}$$

or 138.3 dB.

The power gain is found from

$$A_p \equiv \frac{P_L}{P_I} = \frac{v_L i_o}{v_{i1} i_i}$$

$$= A_v A_i = 818 \times 8.18 \times 10^6 = 66.9 \times 10^8 \text{ W/W}$$

or 98.3 dB. Note that

$$A_p(\text{dB}) = \frac{1}{2}[A_v(\text{dB}) + A_i(\text{dB})]$$

A few comments on the cascade amplifier in the above example are in order. To avoid losing voltage signal strength at the amplifier input, the first stage is designed to have an input resistance much larger than the source resistance (1 MΩ ≫ 100 kΩ). The trade-off appears to be a moderate voltage gain (10 V/V). The second stage realizes the bulk of the required voltage gain. The third and final, or output, stage functions as a buffer amplifier, providing a relatively large input resistance and an output resistance much lower than R_L (10 Ω ≪ 100 Ω). In the following exercises, observe that when finding the gain of an amplifier stage in a cascade amplifier, the loading effect of the succeeding amplifier stage must be taken into account as in the above example.

EXERCISES

1.15 What would the overall voltage gain of the cascade amplifier in Example 1.3 be without stage 3 (i.e., with the load resistance connected to the output of the second stage)?
Ans. 81.8 V/V; a decrease by a factor of 9.

1.16 For the cascade amplifier of Example 1.3, let v_s be 1 mV. Find v_{i1}, v_{i2}, v_{i3}, and v_L.
Ans. 0.91 mV; 9 mV; 818 mV; 744 mV

1.17 (a) Model the three-stage amplifier of Example 1.3 (without the source and load), using the voltage amplifier model of Fig. 1.16(a). What are the values of R_i, A_{vo}, and R_o?
(b) If R_L varies in the range 10 Ω to 1000 Ω, find the corresponding range of the overall voltage gain, v_o/v_s.
Ans. 1 MΩ, 900 V/V, 10 Ω; 409 V/V to 810 V/V

1.5.3 Other Amplifier Types

In the design of an electronic system, the signal of interest—whether at the system input, at an intermediate stage, or at the output—can be either a voltage or a current. For instance, some transducers have a source resistance much larger than the amplifier's input resistance and can therefore be more appropriately modeled as current sources. Similarly, there are applications in which the amplifier output current rather than the voltage is of interest. Thus, although it is the most popular, the voltage amplifier considered above is just one of four possible amplifier types. The other three are the current amplifier, the transconductance amplifier, and the transresistance amplifier. Table 1.1 shows the four amplifier types, their circuit models, the definition of their gain parameters, and the ideal values of their input and output resistances.

1.5.4 Relationships between the Four Amplifier Models

Although for a given amplifier a particular one of the four models in Table 1.1 is most preferable, *any of the four can be used to model any amplifier*. In fact, simple relationships can be derived to relate the parameters of the various models. For instance, the open-circuit voltage gain A_{vo} can be related to the short-circuit current gain A_{is} as follows: The open-circuit output voltage given by the voltage amplifier model of Table 1.1 is $A_{vo}v_i$. The current amplifier model in the same table gives an open-circuit output voltage of $A_{is}i_iR_o$. Equating these two values and noting that $i_i = v_i/R_i$ gives

$$A_{vo} = A_{is}\left(\frac{R_o}{R_i}\right) \tag{1.14}$$

Similarly, we can show that

$$A_{vo} = G_m R_o \tag{1.15}$$

Table 1.1 The Four Amplifier Types

Type	Circuit Model	Gain Parameter	Ideal Characteristics
Voltage Amplifier	i_o, R_o, $+ v_i -$, R_i, $A_{vo}v_i$, $+ v_o -$	Open-Circuit Voltage Gain $A_{vo} \equiv \left.\frac{v_o}{v_i}\right\|_{i_o=0}$ (V/V)	$R_i = \infty$ $R_o = 0$
Current Amplifier	i_i, i_o, R_i, $A_{is}i_i$, R_o, $+ v_o -$	Short-Circuit Current Gain $A_{is} \equiv \left.\frac{i_o}{i_i}\right\|_{v_o=0}$ (A/A)	$R_i = 0$ $R_o = \infty$
Transconductance Amplifier	i_o, $+ v_i -$, R_i, $G_m v_i$, R_o, $+ v_o -$	Short-Circuit Transconductance $G_m \equiv \left.\frac{i_o}{v_i}\right\|_{v_o=0}$ (A/V)	$R_i = \infty$ $R_o = \infty$
Transresistance Amplifier	i_i, R_o, i_o, R_i, $R_m i_i$, $+ v_o -$	Open-Circuit Transresistance $R_m \equiv \left.\frac{v_o}{i_i}\right\|_{i_o=0}$ (V/A)	$R_i = 0$ $R_o = 0$

and

$$A_{vo} = \frac{R_m}{R_i} \tag{1.16}$$

The expressions in Eqs. (1.14) to (1.16) can be used to relate any two of the gain parameters A_{vo}, A_{is}, G_m, and R_m.

Video Example VE 1.3

A 10-mV signal source having an internal resistance of 100 kΩ is connected to an amplifier for which the input resistance is 10 kΩ, the open-circuit voltage gain is 1000 V/V, and the output resistance is 1 kΩ. The amplifier is overall connected in turn to a 100-Ω load.

(a) What overall voltage gain results as measured from the source signal voltage to the load? Where did all the gain go? What would the overall gain be if the source was connected directly to the load? What is the ratio of these two gains? This ratio is a useful measure of the benefit the amplifier brings.
(b) Now instead, replace the source by its Norton equivalent and the amplifier with the equivalent current amplifier from Table 1.1. What is the current gain, i_o/i_s? Show that it is the same as would be computed using the voltage amplifier model.


VE 1.3

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: **1.50**

1.5.5 Determining R_i and R_o

From the amplifier circuit models given in Table 1.1, we observe that the input resistance R_i of the amplifier can be determined by applying an input voltage v_i and measuring (or calculating) the input current i_i; that is, $R_i = v_i/i_i$. The output resistance is found as the ratio of the open-circuit output voltage to the short-circuit output current. Alternatively, the output resistance can be found by eliminating the input signal source (then i_i and v_i will both be zero) and applying a voltage signal v_x to the output of the amplifier, as shown in Fig. 1.18. If we denote the current drawn from v_x *into* the output terminals as i_x (note that i_x is opposite in direction to i_o), then $R_o = v_x/i_x$. Although these techniques are conceptually correct, in actual practice more refined methods are employed in measuring R_i and R_o.

1.5.6 Unilateral Models

The amplifier models considered above are **unilateral**; that is, signal flow is unidirectional, from input to output. Whereas the unilateral model suggests that an amplifier's input current and voltage are completely independent of what is connected to the output, this may not be the case. For example, unintended coupling may allow portions of signals at the amplifier output to appear at its input. We shall not pursue this point further at this time except to mention that more complete models for linear two-port networks are given in Appendix C. Also, in later chapters, we will find it necessary in certain cases to augment the models of Table 1.1 to take into account the nonunilateral nature of some transistor amplifiers.

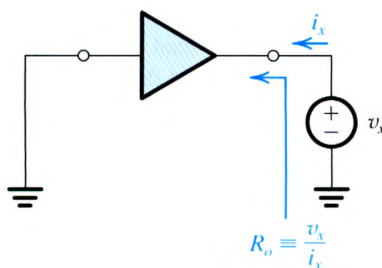


Figure 1.18 Determining the output resistance.

Example 1.4

The **bipolar junction transistor (BJT)**, which will be studied in Chapter 6, is a three-terminal device that when powered up by a dc source (battery) and operated with small signals can be modeled by the linear circuit shown in Fig. 1.19(a). The three terminals are the **base (B)**, the **emitter (E)**, and the **collector (C)**. The heart of the model is a transconductance amplifier represented by an input resistance between B and E (denoted r_π), a short-circuit transconductance g_m, and an output resistance r_o.

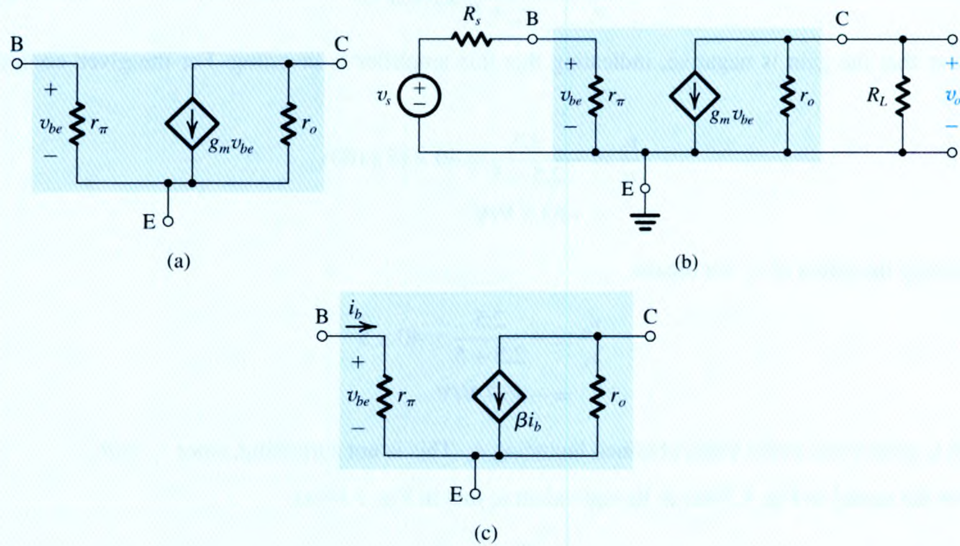


Figure 1.19 **(a)** Small-signal circuit model for a bipolar junction transistor (BJT). **(b)** The BJT connected as an amplifier with the emitter as a common terminal between input and output (called a common-emitter amplifier). **(c)** An alternative small-signal circuit model for the BJT.

(a) With the emitter used as a common terminal between input and output, Fig. 1.19(b) shows a transistor amplifier known as a **common-emitter** or **grounded-emitter** circuit. Derive an expression for the voltage gain v_o/v_s, and evaluate its magnitude for the case $R_s = 5\,\text{k}\Omega$, $r_\pi = 2.5\,\text{k}\Omega$, $g_m = 40\,\text{mA/V}$, $r_o = 100\,\text{k}\Omega$, and $R_L = 5\,\text{k}\Omega$. What would the gain value be if the effect of r_o were neglected?

(b) An alternative model for the transistor in which a current amplifier rather than a transconductance amplifier is used is shown in Fig. 1.19(c). What must the short-circuit current gain β be? Give both an expression and a value.

Solution

(a) Refer to Fig. 1.19(b). We use the voltage-divider rule to determine the fraction of the source signal that appears at the amplifier input as

$$v_{be} = v_s \frac{r_\pi}{r_\pi + R_s} \tag{1.17}$$

Example 1.4 *continued*

Next we determine the output voltage v_o by multiplying the current $(g_m v_{be})$ by the resistance $(R_L \| r_o)$,

$$v_o = -g_m v_{be}(R_L \| r_o) \tag{1.18}$$

Substituting for v_{be} from Eq. (1.17) yields the voltage-gain expression

$$\frac{v_o}{v_s} = -\frac{r_\pi}{r_\pi + R_s} g_m (R_L \| r_o) \tag{1.19}$$

Observe that the gain is negative, indicating that this amplifier is inverting. For the given component values,

$$\begin{aligned}\frac{v_o}{v_s} &= -\frac{2.5}{2.5+5} \times 40 \times (5 \| 100) \\ &= -63.5 \text{ V/V}\end{aligned}$$

Neglecting the effect of r_o, we obtain

$$\begin{aligned}\frac{v_o}{v_s} &\simeq -\frac{2.5}{2.5+5} \times 40 \times 5 \\ &= -66.7 \text{ V/V}\end{aligned}$$

which is quite close to the value obtained including r_o. This is not surprising, since $r_o \gg R_L$.

(b) For the model in Fig. 1.19(c) to be equivalent to that in Fig. 1.19(a),

$$\beta i_b = g_m v_{be}$$

But $i_b = v_{be}/r_\pi$; thus,

$$\beta = g_m r_\pi$$

For the values given,

$$\begin{aligned}\beta &= 40 \text{ mA/V} \times 2.5 \text{ k}\Omega \\ &= 100 \text{ A/A}\end{aligned}$$

EXERCISES

1.18 Consider a current amplifier having the model shown in the second row of Table 1.1. Let the amplifier be fed with a signal current-source i_s having a resistance R_s, and let the output be connected to a load resistance R_L. Show that the overall current gain is given by

$$\frac{i_o}{i_s} = A_{is} \frac{R_s}{R_s + R_i} \frac{R_o}{R_o + R_L}$$

1.19 Consider the transconductance amplifier whose model is shown in the third row of Table 1.1. Let a voltage signal source v_s with a source resistance R_s be connected to the input and a load resistance

R_L be connected to the output. Show that the overall voltage gain is given by

$$\frac{v_o}{v_s} = G_m \frac{R_i}{R_i + R_s}(R_o \| R_L)$$

1.20 Consider a transresistance amplifier having the model shown in the fourth row of Table 1.1. Let the amplifier be fed with a signal current source i_s having a resistance R_s, and let the output be connected to a load resistance R_L. Show that the overall gain is given by

$$\frac{v_o}{i_s} = R_m \frac{R_s}{R_s + R_i} \frac{R_L}{R_L + R_o}$$

1.21 Find the input resistance between terminals B and G in the circuit shown in Fig. E1.21. The voltage v_x is a test voltage with the input resistance R_{in} defined as $R_{in} \equiv v_x/i_x$.

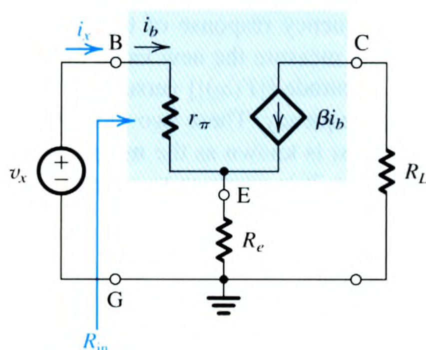


Figure E1.21

Ans. $R_{in} = r_\pi + (\beta + 1)R_e$

1.6 Frequency Response of Amplifiers[2]

From Section 1.2 we know that the input signal to an amplifier can always be expressed as the sum of sinusoidal signals. It follows that an important way to characterize an amplifier is in terms of its response to input sinusoids of different frequencies. Such a characterization of amplifier performance is known as the amplifier frequency response.

1.6.1 Measuring the Amplifier Frequency Response

We begin the discussion of amplifier frequency response by showing how it is measured. Figure 1.20 depicts a linear voltage amplifier fed at its input with a sine-wave signal of amplitude V_i and frequency ω. As the figure indicates, the signal measured at the amplifier

[2]Except for its use in the study of the frequency response of op-amp circuits in Sections 2.5 and 2.7, the material in this section will not be needed in a substantial manner until Chapter 10.

output also is sinusoidal with exactly the same frequency ω. This is important to note: *Whenever a sine-wave signal is applied to a linear circuit, the resulting output is sinusoidal with the same frequency as the input.* In fact, the sine wave is the only signal that does not change shape as it passes through a linear circuit. Observe, however, that the output sinusoid will in general have a different amplitude and will be shifted in phase relative to the input. The ratio of the amplitude of the output sinusoid (V_o) to the amplitude of the input sinusoid (V_i) is the magnitude of the amplifier gain (or transmission) at the test frequency ω. Also, the angle ϕ is the phase of the amplifier transmission at the test frequency ω. If we denote the **amplifier transmission**, or **transfer function** as it is more commonly known, by $T(\omega)$, then

$$|T(\omega)| = \frac{V_o}{V_i}$$

$$\angle T(\omega) = \phi$$

The response of the amplifier to a sinusoid of frequency ω is completely described by $|T(\omega)|$ and $\angle T(\omega)$. Now, to obtain the complete frequency response of the amplifier we simply change the frequency of the input sinusoid and measure the new value for $|T|$ and $\angle T$. The result will be a table and/or graph of gain magnitude [$|T(\omega)|$] versus frequency and a table and/or graph of phase angle [$\angle T(\omega)$] versus frequency. These two plots together constitute the frequency response of the amplifier; the first is known as the **magnitude** or **amplitude response**, and the second is the **phase response**. It is common to express the magnitude of transmission in decibels and thus plot $20 \log |T(\omega)|$ versus frequency.

1.6.2 Amplifier Bandwidth

Figure 1.21 shows the magnitude response of an amplifier. It indicates that the gain is almost constant over a wide frequency range, roughly between ω_1 and ω_2. Signals whose frequencies are below ω_1 or above ω_2 will experience lower gain, with the gain decreasing as we move farther away from ω_1 and ω_2. The band of frequencies over which the gain of the amplifier is almost constant, to within a certain number of decibels (usually 3 dB), is called the **amplifier bandwidth**. Normally the amplifier is designed so that its bandwidth coincides with the spectrum of the signals it is required to amplify. If this were not the case, the amplifier would *distort* the frequency spectrum of the input signal, with different components of the input signal being amplified by different amounts.

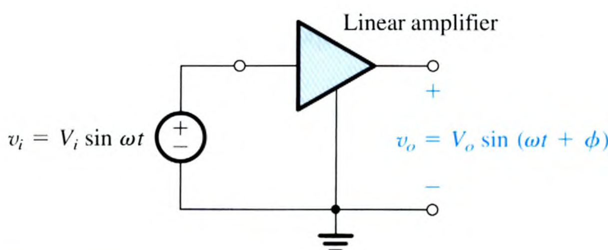


Figure 1.20 Measuring the frequency response of a linear amplifier: At the test frequency, the amplifier gain is characterized by its magnitude (V_o/V_i) and phase ϕ.

1.6.3 Evaluating the Frequency Response of Amplifiers

Having described the method used to measure the frequency response of an amplifier, we now briefly discuss the method for analytically obtaining an expression for the frequency response. This is but a preview of an important subject we will consider at length in Chapter 10.

To evaluate the frequency response of an amplifier, we have to analyze the amplifier equivalent circuit model, taking into account all reactive components.[3] Circuit analysis proceeds in the usual fashion but with inductances and capacitances represented by their reactances. An inductance L has a reactance or impedance $j\omega L$, and a capacitance C has a reactance or impedance $1/j\omega C$ or, equivalently, a susceptance or admittance $j\omega C$. Thus in a *frequency-domain* analysis we deal with impedances and/or admittances. The result of the analysis is the amplifier transfer function $T(\omega)$

$$T(\omega) = \frac{V_o(\omega)}{V_i(\omega)}$$

where $V_i(\omega)$ and $V_o(\omega)$ denote the input and output signals, respectively. $T(\omega)$ is generally a complex function whose magnitude $|T(\omega)|$ gives the magnitude of transmission or the magnitude response of the amplifier. The phase of $T(\omega)$ gives the phase response of the amplifier.

In the analysis of a circuit to determine its frequency response, the algebraic manipulations can be considerably simplified by using the **complex frequency variable** s. In terms of s, the impedance of an inductance L is sL and that of a capacitance C is $1/sC$. Replacing the reactive elements with their impedances and performing standard circuit analysis, we obtain the transfer function $T(s)$ as

$$T(s) \equiv \frac{V_o(s)}{V_i(s)}$$

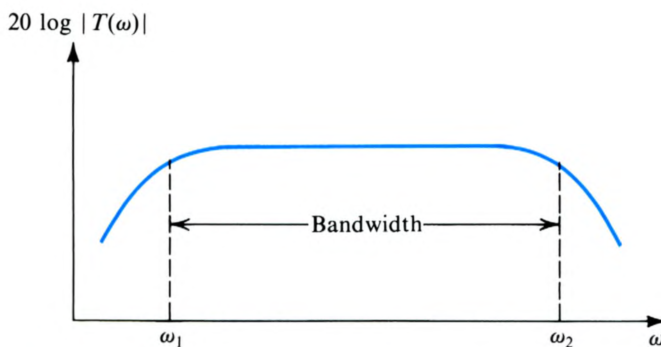


Figure 1.21 Typical magnitude response of an amplifier: $|T(\omega)|$ is the magnitude of the amplifier transfer function—that is, the ratio of the output $V_o(\omega)$ to the input $V_i(\omega)$.

[3] Note that in the models considered in previous sections no reactive components were included. These were simplified models and cannot be used alone to predict the amplifier frequency response.

Subsequently, we replace s by $j\omega$ to determine the transfer function for **physical frequencies**,[4] $T(j\omega)$. Note that $T(j\omega)$ is the same function we called $T(\omega)$ above[5]; the additional j is included in order to emphasize that $T(j\omega)$ is obtained from $T(s)$ by replacing s with $j\omega$.

1.6.4 Single-Time-Constant Networks

In analyzing amplifier circuits to determine their frequency response, we can draw on knowledge of the frequency-response characteristics of single-time-constant (STC) networks. An STC network is composed of, or can be reduced to, one reactive component (inductance or capacitance) and one resistance. Examples are shown in Fig. 1.22. An STC network formed of an inductance L and a resistance R has a time constant $\tau = L/R$. The time constant τ of an STC network composed of a capacitance C and a resistance R is given by $\tau = CR$.

Appendix F presents a study of STC networks and their responses to sinusoidal, step, and pulse inputs. We encourage you to read Appendix F, as you will need an understanding of this material at various points throughout this book. At this point we need the frequency-response results, so we will, in fact, briefly discuss this important topic now.

Most STC networks can be classified into two categories,[6] **low pass (LP)** and **high pass (HP)**, with each of the two categories displaying distinctly different signal responses. As an example, the STC network shown in Fig. 1.22(a) is of the *low-pass* type and that in Fig. 1.22(b) is of the *high-pass* type. To see the reasoning behind this classification, observe that the transfer function of each of these two circuits can be expressed as a voltage-divider ratio, with the divider composed of a resistor and a capacitor. Now, recalling how the impedance of a capacitor varies with frequency ($Z = 1/j\omega C$), it is easy to see that the transmission of the circuit in Fig. 1.22(a) will decrease with frequency and approach zero as ω approaches ∞. Thus the circuit of Fig. 1.22(a) acts as a **low-pass filter**[7]; it passes low-frequency, sine-wave inputs with little or no attenuation (at $\omega = 0$, the transmission is unity) and attenuates high-frequency input sinusoids. The circuit of Fig. 1.22(b) does the opposite; its transmission is unity at $\omega = \infty$ and decreases as ω is reduced, reaching 0 for $\omega = 0$. The latter circuit, therefore, performs as a **high-pass filter**.

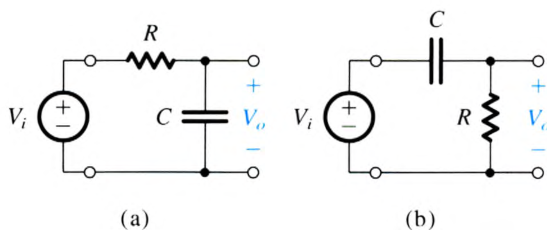


Figure 1.22 Two examples of STC networks: **(a)** a low-pass network and **(b)** a high-pass network.

[4] A "physical frequency" is the frequency, ω, of a sine-wave signal applied at the amplifier input to measure its frequency response.

[5] At this stage, we are using s simply as a shorthand for $j\omega$. We shall not require detailed knowledge of s-plane concepts until Chapter 10. A brief review of s-plane analysis is presented in Appendix F.

[6] An important exception is the **all-pass** STC network.

[7] A filter is a circuit that passes signals in a specified frequency band (the filter passband) and stops or severely attenuates (filters out) signals in another frequency band (the filter stopband). Filters will be studied in Chapter 14.

Table 1.2 summarizes the frequency-response results for STC networks of both types,[8] and sketches of the magnitude and phase responses are given in Figs. 1.23 and 1.24. These frequency-response diagrams are known as **Bode plots**, and the **3-dB frequency** (ω_0) is also known as the **corner frequency**, **break frequency**, or **pole frequency**. We urge you to become familiar with this information and to consult Appendix F if you need further clarification. In particular, you should be able to rapidly determine the time constant τ of an STC circuit. The process is very simple: Set the independent voltage or current source to zero; "grab hold" of the two terminals of the reactive element (capacitor C or inductor L); and determine the equivalent resistance R that appears between these two terminals. The time constant is then CR or L/R.

BODE PLOTS

In the 1930s, while working at Bell Labs, Hendrik Bode devised a simple but accurate method for using linearized asymptotic responses to graph gain and phase shift against frequency on a logarithmic scale. Such gain and phase presentations, together called Bode plots, have enormous importance in the design and analysis of the frequency-dependent behavior of systems large and small.

Table 1.2 Frequency Response of STC Networks

	Low-Pass (LP)	High-Pass (HP)
Transfer Function $T(s)$	$\dfrac{K}{1+(s/\omega_0)}$	$\dfrac{Ks}{s+\omega_0}$
Transfer Function (for physical frequencies) $T(j\omega)$	$\dfrac{K}{1+j(\omega/\omega_0)}$	$\dfrac{K}{1-j(\omega_0/\omega)}$
Magnitude Response $\lvert T(j\omega)\rvert$	$\dfrac{\lvert K\rvert}{\sqrt{1+(\omega/\omega_0)^2}}$	$\dfrac{\lvert K\rvert}{\sqrt{1+(\omega_0/\omega)^2}}$
Phase Response $\angle T(j\omega)$	$-\tan^{-1}(\omega/\omega_0)$	$\tan^{-1}(\omega_0/\omega)$
Transmission at $\omega = 0$ (dc)	K	0
Transmission at $\omega = \infty$	0	K
3-dB Frequency	$\omega_0 = 1/\tau$; $\tau \equiv$ time constant $\tau = CR$ or L/R	
Bode Plots	in Fig. 1.23	in Fig. 1.24

[8]The transfer functions in Table 1.2 are given in general form. For the circuits of Fig. 1.22, $K = 1$ and $\omega_0 = 1/CR$.

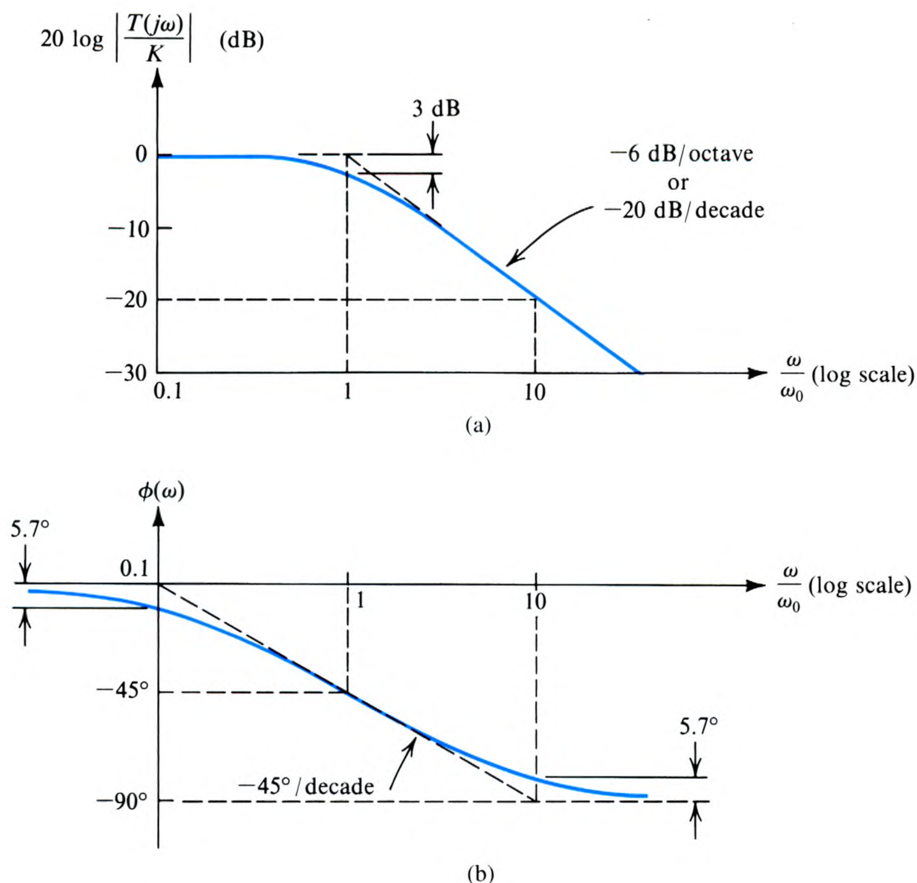


Figure 1.23 **(a)** Magnitude and **(b)** phase response of STC networks of the low-pass type.

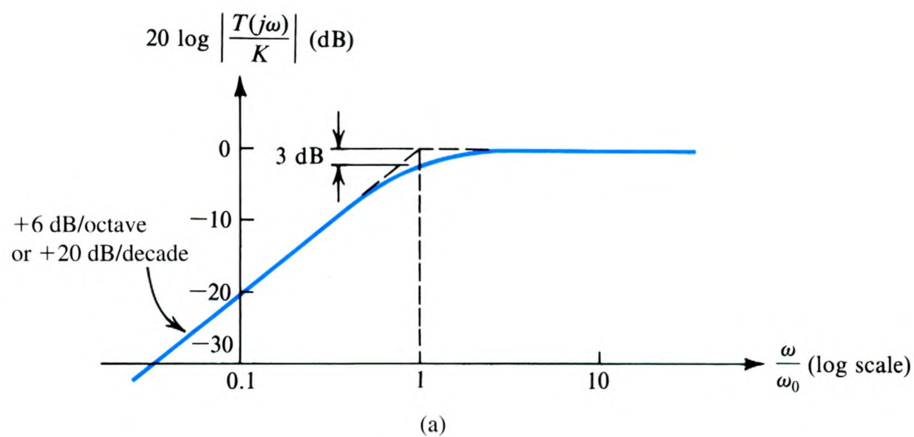


Figure 1.24 **(a)** Magnitude and **(b)** phase response of STC networks of the high-pass type.

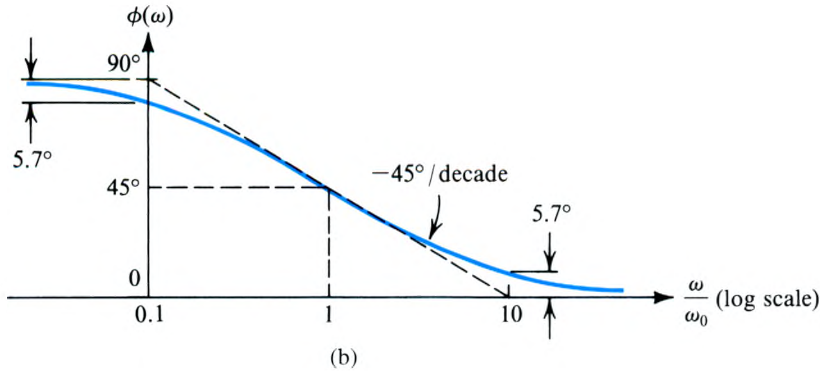


Figure 1.24 *continued*

Video Example VE 1.4

Figure VE1.4 shows a signal source connected to the input of an amplifier. Here R_s is the source resistance, and R_i and C_i are the input resistance and input capacitance, respectively, of the amplifier. Derive an expression for $V_i(s)/V_s(s)$, and show that it is of the low-pass STC type. Find the 3-dB frequency for the case $R_s = 20\text{ k}\Omega$, $R_i = 40\text{ k}\Omega$, and $C_i = 2\text{ pF}$.

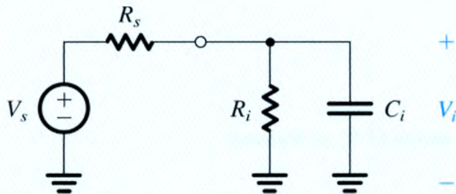


Figure VE1.4 Circuit for Video Example VE 1.4.

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.


VE 1.4

Related end-of-chapter problem: 1.70

Example 1.5

Figure 1.25 shows a voltage amplifier having an input resistance R_i, an input capacitance C_i, a gain factor μ, and an output resistance R_o. The amplifier is fed with a voltage source V_s having a source resistance R_s, and a load of resistance R_L is connected to the output.

Example 1.5 *continued*

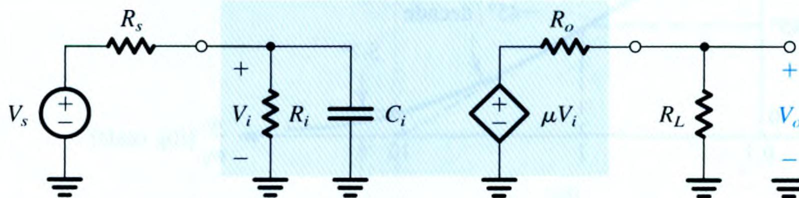


Figure 1.25 Circuit for Example 1.5.

(a) Derive an expression for the amplifier voltage gain V_o/V_s as a function of frequency. From this find expressions for the dc gain and the 3-dB frequency.

(b) Calculate the values of the dc gain, the 3-dB frequency, and the frequency at which the gain becomes 0 dB (i.e., unity) for the case $R_s = 20\ \text{k}\Omega$, $R_i = 100\ \text{k}\Omega$, $C_i = 60\ \text{pF}$, $\mu = 144$ V/V, $R_o = 200\ \Omega$, and $R_L = 1\ \text{k}\Omega$.

(c) Find $v_o(t)$ for each of the following inputs:

(i) $v_i = 0.1 \sin 10^2 t$ V

(ii) $v_i = 0.1 \sin 10^5 t$ V

(iii) $v_i = 0.1 \sin 10^6 t$ V

(iv) $v_i = 0.1 \sin 10^8 t$ V

Solution

(a) Utilizing the voltage-divider rule, we can express V_i in terms of V_s as follows

$$V_i = V_s \frac{Z_i}{Z_i + R_s}$$

where Z_i is the amplifier input impedance. Since Z_i is composed of two parallel elements, it is obviously easier to work in terms of $Y_i = 1/Z_i$. Hence we divide the numerator and denominator by Z_i,

$$V_i = V_s \frac{1}{1 + R_s Y_i} = V_s \frac{1}{1 + R_s[(1/R_i) + sC_i]}$$

$$\Rightarrow \frac{V_i}{V_s} = \frac{1}{1 + (R_s/R_i) + sC_i R_s}$$

This expression can be put in the standard form for a low-pass STC network (see the top line of Table 1.2) by extracting $[1 + (R_s/R_i)]$ from the denominator; thus we have

$$\frac{V_i}{V_s} = \frac{1}{1 + (R_s/R_i)} \frac{1}{1 + sC_i[(R_s R_i)/(R_s + R_i)]} \tag{1.20}$$

At the output side of the amplifier we can use the voltage-divider rule to write

$$V_o = \mu V_i \frac{R_L}{R_L + R_o}$$

This equation can be combined with Eq. (1.20) to obtain the amplifier transfer function

$$\frac{V_o}{V_s} = \mu \frac{1}{1 + (R_s/R_i)} \frac{1}{1 + (R_o/R_L)} \frac{1}{1 + sC_i[(R_s R_i)/(R_s + R_i)]} \tag{1.21}$$

We note that only the last factor in this expression is new (compared with the expression derived in the last section). This factor is a result of the input capacitance C_i, with the time constant being

$$\begin{aligned} \tau &= C_i \frac{R_s R_i}{R_s + R_i} \\ &= C_i (R_s \,\|\, R_i) \end{aligned} \tag{1.22}$$

We could have obtained this result by inspection: From Fig. 1.25 we see that the input circuit is an STC network and that its time constant can be found by reducing V_s to zero, with the result that the resistance seen by C_i is R_i in parallel with R_s. The transfer function in Eq. (1.21) is of the form $K/(1 + (s/\omega_o))$, which corresponds to a low-pass STC network. The dc gain is found as

$$K \equiv \frac{V_o}{V_s}(s = 0) = \mu \frac{1}{1 + (R_s/R_i)} \frac{1}{1 + (R_o/R_L)} \tag{1.23}$$

The 3-dB frequency ω_0 can be found from

$$\omega_0 = \frac{1}{\tau} = \frac{1}{C_i(R_s \,\|\, R_i)} \tag{1.24}$$

Since the frequency response of this amplifier is of the low-pass STC type, the Bode plots for the gain magnitude and phase will take the form shown in Fig. 1.23, where K is given by Eq. (1.23) and ω_0 is given by Eq. (1.24).

(b) Substituting the numerical values given into Eq. (1.23) results in

$$K = 144 \frac{1}{1 + (20/100)} \frac{1}{1 + (200/1000)} = 100 \text{ V/V}$$

Thus the amplifier has a dc gain of 40 dB. Substituting the numerical values into Eq. (1.24) gives the 3-dB frequency

$$\begin{aligned} \omega_0 &= \frac{1}{60 \text{ pF} \times (20 \text{ k}\Omega \,\|\, 100 \text{ k}\Omega)} \\ &= \frac{1}{60 \times 10^{-12} \times (20 \times 100/(20 + 100)) \times 10^3} = 10^6 \text{ rad/s} \end{aligned}$$

Thus,

$$f_0 = \frac{10^6}{2\pi} = 159.2 \text{ kHz}$$

Example 1.5 *continued*

Since the gain falls off at the rate of –20 dB/decade, starting at ω_0 (see Fig. 1.23a) the gain will reach 0 dB in two decades (a factor of 100); thus we have

$$\text{Unity-gain frequency} = 100 \times \omega_0 = 10^8 \text{ rad/s or } 15.92 \text{ MHz}$$

(c) To find $v_o(t)$ we need to determine the gain magnitude and phase at $10^2, 10^5, 10^6$, and 10^8 rad/s. This can be done either approximately utilizing the Bode plots of Fig. 1.23 or exactly utilizing the expression for the amplifier transfer function,

$$T(j\omega) \equiv \frac{V_o}{V_s}(j\omega) = \frac{100}{1 + j(\omega/10^6)}$$

We shall do both:

(i) For $\omega = 10^2$ rad/s, which is $(\omega_0/10^4)$, the Bode plots of Fig. 1.23 suggest that $|T| = K = 100$ and $\phi = 0°$. The transfer function expression gives $|T| \simeq 100$ and $\phi = -\tan^{-1} 10^{-4} \simeq 0°$. Thus,

$$v_o(t) = 10 \sin 10^2 t, \text{ V}$$

(ii) For $\omega = 10^5$ rad/s, which is $(\omega_0/10)$, the Bode plots of Fig. 1.23 suggest that $|T| \simeq K = 100$ and $\phi = -5.7°$. The transfer function expression gives $|T| = 99.5$ and $\phi = -\tan^{-1} 0.1 = -5.7°$. Thus,

$$v_o(t) = 9.95 \sin(10^5 t - 5.7°), \text{ V}$$

(iii) For $\omega = 10^6$ rad/s $= \omega_0$, $|T| = 100/\sqrt{2} = 70.7$ V/V or 37 dB and $\phi = -45°$. Thus,

$$v_o(t) = 7.07 \sin(10^6 t - 45°), \text{ V}$$

(iv) For $\omega = 10^8$ rad/s, which is $(100\,\omega_0)$, the Bode plots suggest that $|T| = 1$ and $\phi = -90°$. The transfer function expression gives $|T| \simeq 1$ and $\phi = -\tan^{-1} 100 = -89.4°$. Thus,

$$v_o(t) = 0.1 \sin(10^8 t - 89.4°), \text{ V}$$

1.6.5 Classification of Amplifiers Based on Frequency Response

Amplifiers can be classified based on the shape of their magnitude-response curve. Figure 1.26 shows typical frequency-response curves for various amplifier types. In Fig. 1.26(a) the gain remains constant over a wide frequency range, but falls off at low and high frequencies. This type of frequency response is common in audio amplifiers since the dc and very-low-frequency components of an audio signal are imperceptible to the human ear.

As we will see in later chapters, **internal capacitances** in the device (a transistor) cause the falloff of gain at high frequencies, just as C_i did in the circuit of Example 1.5. On the other hand, the falloff of gain at low frequencies is usually caused by **coupling capacitors**

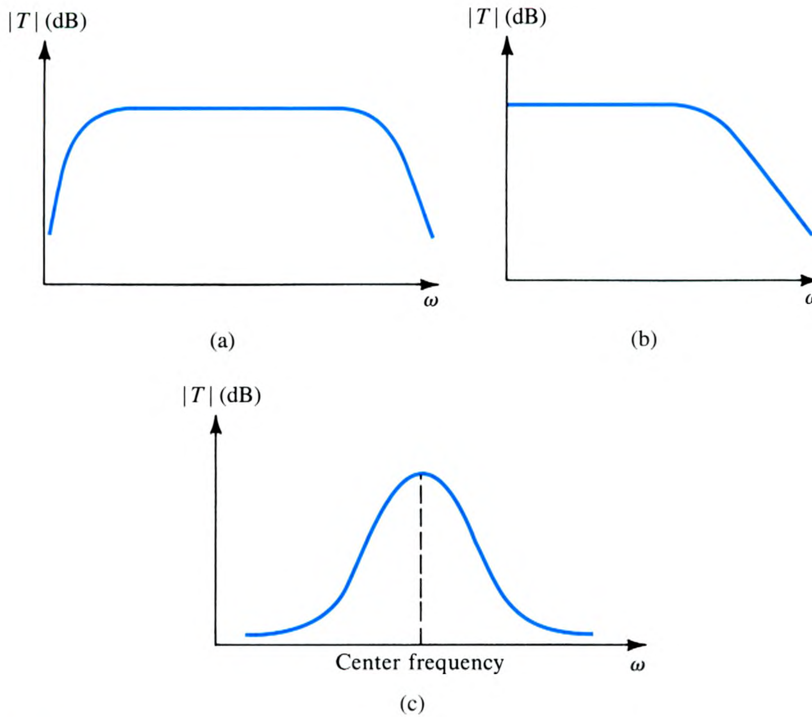


Figure 1.26 Frequency response for **(a)** a capacitively coupled amplifier, **(b)** a direct-coupled amplifier, and **(c)** a tuned or bandpass amplifier.

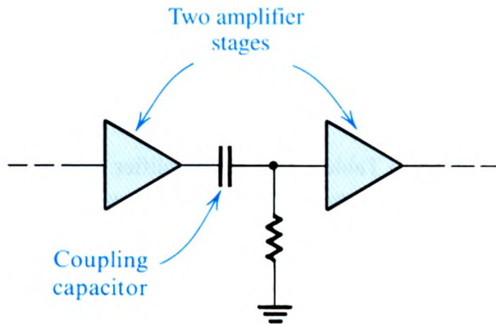


Figure 1.27 Use of a capacitor to couple amplifier stages.

used to connect one amplifier stage to another, as indicated in Fig. 1.27. This practice is usually adopted to simplify the design process of the different stages. Also, some sensors require large dc voltages for proper operation. Because this may damage the following amplifier, a coupling capacitor is inserted between them. The coupling capacitors are usually chosen quite large (a fraction of a microfarad to a few tens of microfarads) so that their reactance (impedance) is small at the frequencies of interest. Nevertheless, at sufficiently low frequencies the reactance of a coupling capacitor will become large enough to cause part of the signal being coupled to appear as a voltage drop across the coupling capacitor, thus not reaching the subsequent stage. Coupling capacitors will thus cause loss of gain at low frequencies and cause the gain to be zero at dc. This is not at all surprising, since from

Fig. 1.27 we observe that the coupling capacitor, acting together with the input resistance of the subsequent stage, forms a high-pass STC circuit. It is the frequency response of this high-pass circuit that accounts for the shape of the amplifier frequency response in Fig. 1.26(a) at the low-frequency end.

There are many applications in which it is important that the amplifier maintain its gain at low frequencies down to dc. This is the case when the signal of interest is nearly constant, for example, the signal produced by a temperature sensor in a refrigerator. Furthermore, monolithic integrated-circuit (IC) technology does not allow the fabrication of large coupling capacitors. Thus IC amplifiers are usually designed as **directly coupled** or **dc amplifiers** (as opposed to **capacitively coupled**, or **ac amplifiers**). Figure 1.26(b) shows the frequency response of a dc amplifier. Such a frequency response characterizes what is referred to as a **low-pass amplifier**.

In a number of applications, such as in the design of radio and TV receivers, the need arises for an amplifier whose frequency response peaks around a certain frequency (called the **center frequency**) and falls off on both sides of this frequency, as shown in Fig. 1.26(c). Amplifiers with such a response are called **tuned amplifiers**, **bandpass amplifiers**, or **bandpass filters**. A tuned amplifier forms the heart of the front-end or tuner of a communication receiver; by adjusting its center frequency to coincide with the frequency of a desired communications channel (e.g., a radio station), the signal of this particular channel can be received while those of other channels are attenuated or filtered out. Tuned amplifiers are also useful for the sinusoidal signals that are commonly used as clocks in electronic systems while attenuating noise and interference at other frequencies.

EXERCISES

1.22 Consider a voltage amplifier having a frequency response of the low-pass STC type with a dc gain of 60 dB and a 3-dB frequency of 1000 Hz. Find the gain in dB at $f = 10$ Hz, 10 kHz, 100 kHz, and 1 MHz.

Ans. 60 dB; 40 dB; 20 dB; 0 dB

D1.23 Consider a transconductance amplifier having the model shown in Table 1.1. If the amplifier load consists of a resistance R_L in parallel with a capacitance C_L, convince yourself that the voltage transfer function realized, V_o/V_i, is of the low-pass STC type. For example, this may be a suitable model when a transistor is driving an ultrasonic transducer. Consider the case where the transducer is a load with $R_L = 1$ kΩ and $C_L = 4.5$ nF, and find the largest value of R_o for which a 3-dB bandwidth of at least 40 kHz is provided. With this value of R_o, find the minimum value of G_m required to ensure a dc gain of at least 40 dB.

Ans. 7.6 kΩ; 113 mA/V

D1.24 Consider the situation illustrated in Fig. 1.27. Let the output resistance of the first voltage amplifier be 1 kΩ and the input resistance of the second voltage amplifier (including the resistor shown) be 9 kΩ. The resulting equivalent circuit is shown in Fig. E1.24. Convince yourself that V_2/V_s is a high-pass STC function. What is the smallest value for C that will ensure that the 3-dB frequency is not higher than 100 Hz?

Ans. 0.16 μF

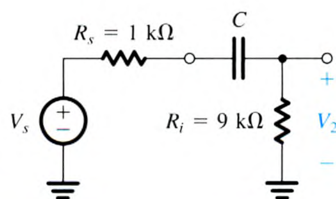


Figure E1.24

Summary

- An electrical signal source can be represented in either the Thévenin form (a voltage source v_s in series with a source resistance R_s) or the Norton form (a current source i_s in parallel with a source resistance R_s). The Thévenin voltage v_s is the open-circuit voltage between the source terminals; the Norton current i_s is equal to the short-circuit current between the source terminals. For the two representations to be equivalent, v_s and $R_s i_s$ must be equal.
- A signal can be represented either by its waveform versus time or as the sum of sinusoids. The latter representation is known as the frequency spectrum of the signal.
- The sine-wave signal is completely characterized by its peak value (or rms value, which is the peak/$\sqrt{2}$), its frequency (ω in rad/s or f in Hz; $\omega = 2\pi f$ and $f = 1/T$, where T is the period in seconds), and its phase with respect to an arbitrary reference time.
- Analog signals have magnitudes that can assume any value. Electronic circuits that process analog signals are called analog circuits. Sampling the magnitude of an analog signal at discrete instants of time and representing each signal sample by a number results in a digital signal. Digital signals are processed by digital circuits.
- The simplest digital signals are obtained when the binary system is used. An individual digital signal then assumes one of only two possible values: low and high (say, 0 V and +1.8 V), corresponding to logic 0 and logic 1, respectively.
- An analog-to-digital converter (ADC) provides at its output the digits of the binary number representing the analog signal sample applied to its input. The output digital signal can then be processed using digital circuits. Refer to Fig. 1.10 and Eq. (1.3).
- The transfer characteristic, v_O versus v_I, of a linear amplifier is a straight line with a slope equal to the voltage gain. Refer to Fig. 1.12.
- Amplifiers increase the signal power and thus require dc power supplies for their operation.
- The amplifier voltage gain can be expressed as a ratio A_v in V/V or in decibels, 20 log $|A_v|$, dB. Similarly, for current gain: A_i A/A or 20 log $|A_i|$, dB. For power gain: A_p W/W or 10 log A_p, dB.
- Depending on the signal to be amplified (voltage or current) and on the desired form of output signal (voltage or current), there are four basic amplifier types: voltage, current, transconductance, and transresistance amplifiers. For the circuit models and ideal characteristics of these four amplifier types, refer to Table 1.1. A given amplifier can be modeled by any one of the four models, in which case their parameters are related by the formulas in Eqs. (1.14) to (1.16).
- The sinusoid is the only signal whose waveform is unchanged through a linear circuit. Sinusoidal signals are used to measure the frequency response of amplifiers.
- The transfer function $T(s) \equiv V_o(s)/V_i(s)$ of a voltage amplifier can be determined from circuit analysis. Substituting $s = j\omega$ gives $T(j\omega)$, whose magnitude $|T(j\omega)|$ is the magnitude response, and whose phase $\phi(\omega)$ is the phase response, of the amplifier.
- Amplifiers are classified according to the shape of their frequency response, $|T(j\omega)|$. Refer to Fig. 1.26.
- Single-time-constant (STC) networks are those networks that are composed of, or can be reduced to, one reactive component (L or C) and one resistance (R). The time constant τ is either L/R or CR.
- STC networks can be classified into two categories: low pass (LP) and high pass (HP). LP networks pass dc and low frequencies and attenuate high frequencies. The opposite is true for HP networks.
- The gain of an LP (HP) STC circuit drops by 3 dB below the zero-frequency (infinite-frequency) value at a frequency $\omega_0 = 1/\tau$. At high frequencies (low frequencies) the gain falls off at the rate of 6 dB/octave or 20 dB/decade. Refer to Table 1.2 and Figs. 1.23 and 1.24. Further details are given in Appendices E and F.

Circuit Basics

As a review of the basics of circuit analysis and in order for the readers to gauge their preparedness for the study of electronic circuits, this section presents a number of relevant circuit analysis problems. For a summary of Thévenin's and Norton's theorems, refer to Appendix D. The problems are grouped in appropriate categories.

Resistors and Ohm's Law

1.1 Ohm's law relates V, I, and R for a resistor. For each of the situations following, find the missing item:

(a) $R = 1\ \text{k}\Omega, V = 0.5\ \text{V}$
(b) $V = 2\ \text{V}, I = 1\ \text{mA}$
(c) $R = 20\ \text{k}\Omega,\ I = 0.1\ \text{mA}$
(d) $R = 100\ \Omega, V = 5\ \text{V}$

Note: Volts, milliamps, and kilohms constitute a consistent set of units.

1.2 Measurements taken on various resistors are shown below. For each, calculate the power dissipated in the resistor and the power rating necessary for safe operation using standard components with power ratings of 1/8 W, 1/4 W, 1/2 W, 1 W, or 2 W:

(a) 1 kΩ conducting 20 mA
(b) 1 kΩ conducting 40 mA
(c) 100 kΩ conducting 1 mA
(d) 10 kΩ conducting 4 mA
(e) 1 kΩ dropping 20 V
(f) 1 kΩ dropping 11 V

1.3 Ohm's law and the power law for a resistor relate V, I, R, and P, making only two variables independent. For each pair identified below, find the other two:

(a) $R = 1\ \text{k}\Omega,\ I = 2\ \text{mA}$
(b) $V = 1\ \text{V}, I = 20\ \text{mA}$
(c) $V = 1\ \text{V}, P = 100\ \text{mW}$
(d) $I = 0.1\ \text{mA}, P = 2\ \text{mW}$
(e) $R = 1\ \text{k}\Omega,\ P = 0.1\ \text{W}$

Combining Resistors

1.4 You have available only 5-kΩ resistors, but you have as many as you would like. Using series and parallel combinations of them, realize the following resistor values: 20 kΩ, 1.67 kΩ, 12.5 kΩ, 23.75 kΩ.

1.5 In the analysis and test of electronic circuits, it is often useful to connect one resistor in parallel with another to obtain a nonstandard value, one which is smaller than the smaller of the two resistors. Often, particularly during circuit testing, one resistor is already installed, in which case the second, when connected in parallel, is said to "shunt" the first. If the original resistor is 10 kΩ, what is the value of the shunting resistor needed to reduce the combined value by 1%, 5%, 10%, and 50%? What is the result of shunting a 10-kΩ resistor by 1 MΩ? By 100 kΩ? By 10 kΩ?

Voltage Dividers

1.6 Figure P1.6(a) shows a two-resistor voltage divider. Its function is to generate a voltage V_O (smaller than the power-supply voltage V_{DD}) at its output node X. The circuit looking back at node X is equivalent to that shown in Fig. P1.6(b). Observe that this is the Thévenin equivalent of the voltage-divider circuit. Find expressions for V_O and R_O.

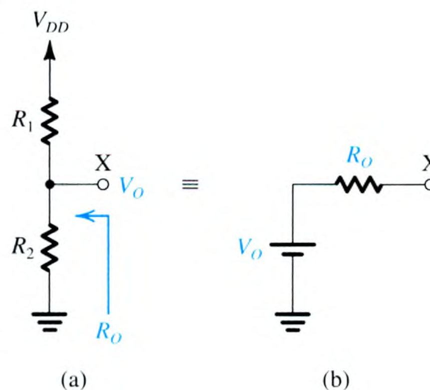


Figure P1.6

1.7 A two-resistor voltage divider employing a 2-kΩ and a 1-kΩ resistor is connected to a 3-V ground-referenced power supply to provide a 2-V voltage. Sketch the circuit. Assuming exact-valued resistors, what output voltage (measured to ground) and equivalent output resistance result? If the resistors used are not ideal but have a ±5% manufacturing tolerance, what are the extreme output voltages and resistances that can result?

problems with blue numbers are considered essential; * = difficult problem; ** = more difficult; *** = very challenging
SIM = simulation; **D** = design problem; ● = see related video example

D 1.8 You are given three resistors, each of 10 kΩ, and a 9-V battery whose negative terminal is connected to ground. With a voltage divider using some or all of your resistors, how many positive-voltage sources of magnitude less than 9 V can you design? List them in order, smallest first. What is the output resistance (i.e., the Thévenin resistance) of each?

D *1.9 Two resistors, with nominal values of 470 Ω and 1 kΩ, are used in a voltage divider with a +3-V supply to create a nominal +1-V output. Assuming the resistor values to be exact, what is the actual output voltage produced? Which resistor must be shunted (paralleled) by what third resistor to create a voltage-divider output of 1.00 V? If an output resistance of exactly 333 Ω is also required, what do you suggest?

Current Dividers

1.10 Figure P1.10 shows a two-resistor current divider fed with an ideal current source I. Show that

$$I_1 = \frac{R_2}{R_1 + R_2} I$$

$$I_2 = \frac{R_1}{R_1 + R_2} I$$

and find the voltage V that develops across the current divider.

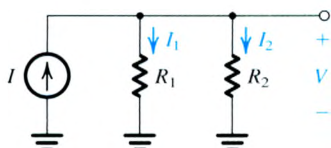


Figure P1.10

1.11 Fig. P1.11 shows a current source driving N resistors in parallel. Show that the current through the kth resistor is

$$I_k = \frac{I/R_k}{\sum_{i=1}^{N} 1/R_i}$$

Show that when $N = 2$, this expression is the same as for a two-resistor current divider.

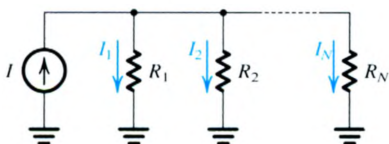


Figure P1.11

D 1.12 Design a simple current divider that will reduce the current provided to a 6-kΩ load to one-quarter of that available from the source.

D 1.13 A designer searches for a simple circuit to provide one-fifth of a signal current I to a load resistance R. Suggest a solution using one resistor. What must its value be? What is the input resistance of the resulting current divider? For a particular value R, the designer discovers that the otherwise-best-available resistor is 10% too high. Suggest two circuit topologies using one additional resistor that will solve this problem. What is the value of the resistor required in each case? What is the input resistance of the current divider in each case?

D 1.14 A current source produces a sinusoidal signal with 0.5 mA peak amplitude. However, the voltage across the source must not exceed 1 V, otherwise the signal becomes distorted. What is the maximum load resistance that may be connected to the source while respecting this limit? If a 10-kΩ load is to be connected, what additional resistor must be connected in parallel to prevent distortion? What is the resulting current signal through the 10-kΩ load?

Thévenin Equivalent Circuits

1.15 For the circuit in Fig. P1.15, find the Thévenin equivalent circuit between terminals (a) 1 and 2, (b) 2 and 3, and (c) 1 and 3.

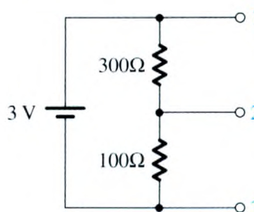


Figure P1.15

1.16 Through repeated application of Thévenin's theorem, find the Thévenin equivalent of the circuit in Fig. P1.16 between node 4 and ground, and hence find the current that flows through a load resistance of 3 kΩ connected between node 4 and ground.

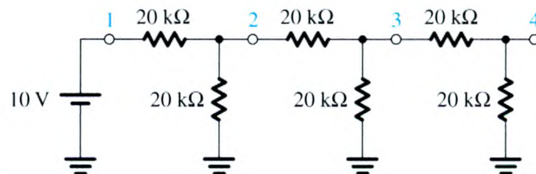


Figure P1.16

CHAPTER 1 PROBLEMS

Circuit Analysis

VE 1.1 **1.17** For the circuit shown in Fig. P1.17, find the current in each of the three resistors and the voltage (with respect to ground) at their common node using two methods:

(a) Loop Equations: Define branch currents I_1 and I_2 in R_1 and R_2, respectively; write two equations; and solve them.

(b) Node Equation: Define the node voltage V at the common node; write a single equation; and solve it.

Which method do you prefer? Why?

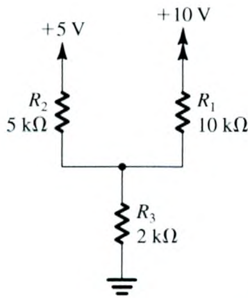


Figure P1.17

1.18 The circuit shown in Fig. P1.18 represents the equivalent circuit of an unbalanced bridge. Such equivalent circuits arise, for example, in pressure sensors where the resisances R_{1-4} vary with pressure and the output signal is observed across terminals 1 and 2. It is required to calculate the current in the detector branch (R_5) and the voltage across it. Although this can be done by using loop and node equations, a much easier approach is possible: Find the Thévenin equivalent of the circuit to the left of node 1 and the Thévenin equivalent of the circuit to the right of node 2. Then solve the resulting simplified circuit.

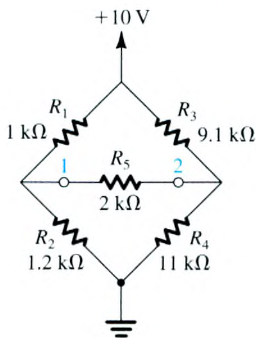


Figure P1.18

1.19 Two sources are connected as shown in Fig. P1.19 below. Use Thévenin equivalent representations to help find the voltage v_o.

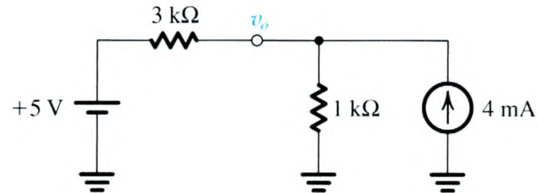


Figure P1.19

1.20 Derive an expression for v_o/v_s for the circuit shown in Fig. P1.20.

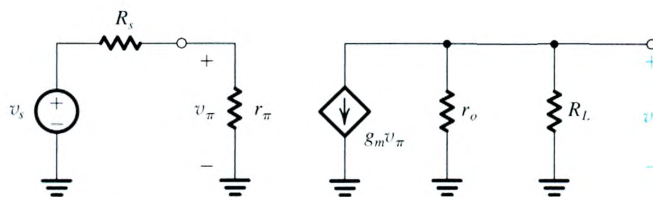


Figure P1.20

AC Circuits

1.21 The periodicity of recurrent waveforms, such as sine waves or square waves, can be completely specified using only one of three possible parameters: radian frequency, ω, in radians per second (rad/s); (conventional) frequency, f, in hertz (Hz); or period, T, in seconds (s). As well, each of the parameters can be specified numerically in one of several ways: using letter prefixes associated with the basic units, using scientific notation, or using some combination of both. Thus, for example, a particular period may be specified as 100 ns, 0.1 μs, 10^{-1} μs, 10^5 ps, or 1×10^{-7} s. (For the definition of the various prefixes used in electronics, see Appendix J.) For each of the measures listed below, express the trio of terms in scientific notation associated with the basic unit (e.g., 10^{-7} s rather than 10^{-1} μs).

(a) $T = 10^{-4}$ μs
(b) $f = 3$ GHz
(c) $\omega = 6.28 \times 10^4$ rad/s
(d) $T = 10^{-7}$ s
(e) $f = 60$ Hz
(f) $\omega = 100$ krad/s
(g) $f = 270$ MHz

1.22 Find the complex impedance, Z, of each of the following basic circuit elements at 60 Hz, 100 kHz, and 1 GHz:

(a) $R = 1$ kΩ
(b) $C = 10$ nF
(c) $C = 10$ pF
(d) $L = 10$ mH
(e) $L = 1$ μH

1.23 Find the complex impedance at 10 kHz of the following networks:

(a) 1 kΩ in series with 1 nF
(b) 1 kΩ in parallel with 0.01 μF
(c) 10 kΩ in parallel with 100 pF
(d) 100 kΩ in series with 10 mH

***1.24** Write an expression for the complex impedance of a parallel combination of an inductor, L, and a capacitor, C, as a function of frequency.

Find the frequency at which the magnitude of the impedance becomes infinite. What happens to the phase of the impedance at this frequency? What current will the parallel combination draw from an ideal sinusoidal voltage source at this frequency?

Section 1.1: Signals

1.25 Any given signal source provides an open-circuit voltage, v_{oc}, and a short-circuit current, i_{sc}. For the following sources, calculate the internal resistance, R_s; the Norton current, i_s; and the Thévenin voltage, v_s:

(a) $v_{oc} = 3$ V, $i_{sc} = 3$ mA
(b) $v_{oc} = 500$ mV, $i_{sc} = 50$ μA

1.26 A particular signal source produces an output of 100 μA when loaded by a 100-kΩ resistor and 500 μA when loaded by a 10-kΩ resistor. Calculate the Thévenin voltage, Norton current, and source resistance.

1.27 An audio signal source is connected to a speaker. When connected to a 16-Ω speaker, the source delivers 25% less power than when connected to a 32-Ω headphone speaker. What is the source resistance?

1.28 A temperature sensor is specified to provide 2 mV/°C. When connected to a load resistance of 10 kΩ, the output voltage was measured to change by 10 mV, corresponding to a change in temperature of 20°C. What is the source resistance of the sensor?

1.29 Refer to the Thévenin and Norton representations of the signal source (Fig. 1.1). If the current supplied by the source is denoted i_o and the voltage appearing between the source output terminals is denoted v_o, sketch and clearly label v_o versus i_o for $0 \le i_o \le i_s$.

VE 1.2

1.30 Consider voltage sources connected to loads with the values shown below. In each case, find the percentage change in V_L and I_L in response to a 10% increase in the value of R_L. In which cases is it more appropriate to use a Norton equivalent source? In those cases, find the Norton equivalent for $V_s = 1$ V.

(a) $R_s = 5$ kΩ; $R_L = 200$ kΩ
(b) $R_s = 5$ Ω; $R_L = 50$ Ω
(c) $R_s = 2$ kΩ; $R_L = 100$ Ω
(d) $R_s = 150$ Ω; $R_L = 16$ Ω

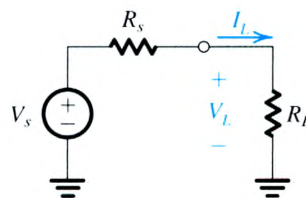


Figure P1.30

1.31 The connection of a signal source to an associated signal processor or amplifier generally involves some degree of signal loss as measured at the processor or amplifier input. Considering the two signal-source representations shown in Fig. 1.1, provide two sketches showing each signal-source representation connected to the input terminals (and corresponding input resistance) of a signal processor. What signal-processor input resistance will result in 95% of the open-circuit voltage being delivered to the processor? What input resistance will result in 95% of the short-circuit signal current entering the processor?

Section 1.2: Frequency Spectrum of Signals

1.32 To familiarize yourself with typical values of angular frequency ω, conventional frequency f, and period T, complete the entries in the following table:

Case	ω (rad/s)	f (Hz)	T (s)
a		5×10^9	
b	2×10^9		
c			1×10^{-10}
d		60	
e	6.28×10^4		
f			1×10^{-5}

1.33 For the following peak or rms values of some important sine waves, calculate the corresponding other value:

(a) 117 V rms, a household-power voltage in North America
(b) 33.9 V peak, a somewhat common voltage in rectifier circuits
(c) 220 V rms, a household-power voltage in parts of Europe
(d) 220 kV rms, a high-voltage transmission-line voltage in North America

1.34 Give expressions for the sine-wave voltage signals having:

(a) 3-V peak amplitude and 20-kHz frequency
(b) 120-V rms and 60-Hz frequency
(c) 0.2-V peak-to-peak and 10^8-rad/s frequency
(d) 100-mV peak and 1-ns period

1.35 Characterize a symmetrical square-wave wave with peak-to-peak amplitude 2V and period 0.5 ms. Sketch the waveform. What is its average value? Its lowest value? Its highest value? Its frequency? Write an expression for it in terms of its sinusoidal components.

1.36 Measurements taken of a square-wave signal using a frequency-selective voltmeter (called a spectrum analyzer) show its spectrum to contain adjacent components (spectral lines) at 98 kHz and 126 kHz of amplitudes 63 mV and 49 mV, respectively. What would direct measurement of the fundamental show its frequency and amplitude to be? What is the rms value of the fundamental? What are the peak-to-peak amplitude and period of the originating square wave?

1.37 A symmetrical square wave with peak amplitude $\hat{V}$ and zero average is to be approximated by its first five (lowest frequency) sinusoidal components. Compare the rms value of the square wave to that of its approximation. What is the percentage difference?

1.38 Find the amplitude of a symmetrical square wave of period T that provides the same power as a sine wave of peak amplitude $\hat{V}$ and the same frequency. Does this result depend on equality of the frequencies of the two waveforms?

Section 1.3: Analog and Digital Signals

1.39 Give the binary representation of the following decimal numbers: 0, 5, 13, 32, and 63.

1.40 Consider a 4-bit digital word $b_3b_2b_1b_0$ in a format called signed-magnitude, in which the most significant bit, b_3, is interpreted as a sign bit—0 for positive and 1 for negative values. List the values that can be represented by this scheme. What is peculiar about the representation of zero? For a particular analog-to-digital converter (ADC), each change in b_0 corresponds to a 0.5-V change in the analog input. What is the full range of the analog signal that can be represented? What signed-magnitude digital code results for an input of +2.5 V? For −3.0 V? For +2.7 V? For −2.8 V?

1.41 Consider an N-bit DAC whose output varies from 0 to V_{FS} (where the subscript FS denotes "full-scale").

(a) Show that a change in the least significant bit (LSB) induces a change of $V_{FS}/(2^N - 1)$ in the output. This is the resolution of the converter.
(b) Convince yourself that the DAC can produce any desired output voltage between 0 and V_{FS} with at most $V_{FS}/2(2^N - 1)$ error (i.e., one-half the resolution). This is called the quantization error of the converter.
(c) For $V_{FS} = 5$ V, how many bits are required to obtain a resolution of 2 mV or better? What is the actual resolution obtained? What is the resulting quantization error?

1.42 Figure P1.42 shows the circuit of an N-bit DAC. Each of the N bits of the digital word to be converted controls one of the switches. When the bit is 0, the switch is in the position labeled 0; when the bit is 1, the switch is in the position labeled 1. The analog output is the current i_O. V_{ref} is a constant reference voltage.

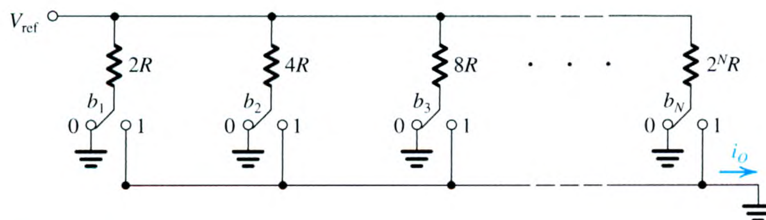


Figure P1.42

(a) Show that

$$i_O = \frac{V_{ref}}{R}\left(\frac{b_1}{2^1} + \frac{b_2}{2^2} + \cdots + \frac{b_N}{2^N}\right)$$

(b) Which bit is the LSB? Which is the MSB?

(c) For $V_{ref} = 10$ V, $R = 10$ kΩ, and $N = 8$, find the maximum value of i_O obtained. What is the change in i_O resulting from the LSB changing from 0 to 1?

1.43 An audio signal is sampled at 44.1 kHz. Each sample is represented by 16 bits. What is the speed of this system in bits per second?

1.44 Each pixel in a 10-megapixel image is represented by 8 bits for the intensity of red, 8 bits for green, and 8 bits for blue. How many such images can be stored in 16 Gbits of memory?

Section 1.4: Amplifiers

1.45 Various amplifier and load combinations are measured as listed below using rms values. For each, find the voltage, current, and power gains (A_v, A_i, and A_p, respectively) both as ratios and in dB:

(a) $v_I = 100$ mV, $i_I = 100$ μA, $v_O = 10$ V, $R_L = 100$ Ω
(b) $v_I = 10$ μV, $i_I = 100$ nA, $v_O = 1$ V, $R_L = 10$ kΩ
(c) $v_I = 1$ V, $i_I = 1$ mA, $v_O = 5$ V, $R_L = 10$ Ω

1.46 An amplifier operating from ±3-V supplies provides a 2.2-V peak sine wave across a 100-Ω load when provided with a 0.2-V peak input from which 1.0 mA peak is drawn. Find the voltage gain, current gain, and power gain expressed as ratios and in decibels. If the amplifier efficiency is 10%, find the supply power, supply current, and amplifier dissipation.

1.47 An amplifier using balanced power supplies is known to saturate for signals extending within 1.0 V of either supply. For linear operation, its gain is 200 V/V. What is the rms value of the largest undistorted sine-wave output available, and input needed, with ±5-V supplies? With ±10-V supplies? With ±15-V supplies?

Section 1.5: Circuit Models for Amplifiers

1.48 Consider the voltage-amplifier circuit model shown in Fig. 1.16(b), in which $A_{vo} = 100$ V/V under the following conditions:

(a) $R_i = 10R_s$, $R_L = 10R_o$
(b) $R_i = R_s$, $R_L = R_o$
(c) $R_i = R_s/10$, $R_L = R_o/10$

Calculate the overall voltage gain v_o/v_s in each case, expressed both directly and in decibels.

1.49 An amplifier with 40 dB of small-signal, open-circuit voltage gain, an input resistance of 1 MΩ, and an output resistance of 100 Ω, drives a load of 500 Ω. What voltage and power gains (expressed in dB) would you expect with the load connected? If the amplifier has a peak output-current limitation of 20 mA, what is the rms value of the largest sine-wave input for which an undistorted output is possible? What is the corresponding output power available?


VE 1.3

1.50 A 10-mV signal source having an internal resistance of 5 kΩ is connected to an amplifier for which the input resistance is 1 kΩ, the open-circuit voltage gain is 100 V/V, and the output resistance is 200 Ω. The amplifier is connected in turn to a 100-Ω load.

(a) What overall voltage gain results as measured from the source internal voltage to the load? Where did all the gain go? What would the gain be if the source was connected directly to the load? What is the ratio of these two gains? This ratio is a useful measure of the benefit the amplifier brings.

(b) Now instead, replace the source by its Norton equivalent and the amplifier with the equivalent current amplifier from Table 1.1. What is the current gain, i_o/i_s? Show that

it is the same as would be computed using the voltage amplifier model.

1.51 A buffer amplifier with a gain of 1 V/V has an input resistance of 1 MΩ and an output resistance of 20 Ω. It is connected between a 1-V, 200-kΩ source and a 100-Ω load. What load voltage results? What are the corresponding voltage, current, and power gains (in dB)?

1.52 Consider the cascade amplifier of Example 1.3. Find the overall voltage gain v_o/v_s obtained when the first and second stages are interchanged. Compare this value with the result in Example 1.3, and comment.

1.53 You are given two amplifiers, A and B, to connect in cascade between a 10-mV, 100-kΩ source and a 100-Ω load. The amplifiers have voltage gain, input resistance, and output resistance as follows: for A, 100 V/V, 100 kΩ, 10 kΩ, respectively; for B, 10 V/V, 10 kΩ, 1 kΩ, respectively. Your problem is to decide how the amplifiers should be connected. To proceed, evaluate the two possible connections between source S and load L, namely, SABL and SBAL. Find the voltage gain for each both as a ratio and in decibels. Which amplifier arrangement is best?

D *1.54 A designer has available voltage amplifiers with an input resistance of 100 kΩ, an output resistance of 1 kΩ, and an open-circuit voltage gain of 15. The signal source has a 50-kΩ resistance and provides a 5-mV rms signal, and it is required to provide a signal of at least 3 V rms to a 200-Ω load. How many amplifier stages are required? What is the output voltage actually obtained?

D *1.55 Design an amplifier that provides 0.5 W of signal power to a 100-Ω load resistance. The signal source provides a 30-mV rms signal and has a resistance of 0.5 MΩ. Three types of voltage-amplifier stages are available:

(a) A high-input-resistance type with $R_i = 1$ MΩ, $A_{vo} = 10$, and $R_o = 10$ kΩ

(b) A high-gain type with $R_i = 10$ kΩ, $A_{vo} = 100$, and $R_o = 1$ kΩ

(c) A low-output-resistance type with $R_i = 10$ kΩ, $A_{vo} = 1$, and $R_o = 20$ Ω

Design a suitable amplifier using a combination of these stages. Your design should utilize the minimum number of stages and should ensure that the signal level is not reduced below 10 mV at any point in the amplifier chain. Find the load voltage and power output realized.

D *1.56 You are required to design a voltage amplifier to be driven from a signal source having a 5-mV peak amplitude and a source resistance of 10 kΩ to supply a peak output of 2 V across a 1-kΩ load.

(a) What is the required voltage gain from the source to the load?

(b) If the peak current available from the source is 0.1 μA, what is the smallest input resistance allowed? For the design with this value of R_i, find the overall current gain and power gain.

(c) If the amplifier power supply limits the peak value of the output open-circuit voltage to 3 V, what is the largest output resistance allowed?

(d) For the design with R_i as in (b) and R_o as in (c), what is the required value of open-circuit voltage gain, i.e., $\left.\frac{v_o}{v_i}\right|_{R_L=\infty}$, of the amplifier?

(e) If, as a possible design option, you are able to increase R_i to the nearest value of the form 1×10^n Ω and to decrease R_o to the nearest value of the form 1×10^m Ω, find (i) the input resistance achievable; (ii) the output resistance achievable; and (iii) the open-circuit voltage gain now required to meet the specifications.

D 1.57 A voltage amplifier with an input resistance of 40 kΩ, an output resistance of 100 Ω, and a gain of 300 V/V is connected between a 10-kΩ source with an open-circuit voltage of 10 mV and a 100-Ω load. For this situation:

(a) What output voltage results?

(b) What is the voltage gain from source to load?

(c) What is the voltage gain from the amplifier input to the load?

(d) If the output voltage across the load is twice that needed and there are signs of amplifier saturation, suggest the location and value of a single resistor that would produce the desired output. Choose an arrangement that would cause minimum disruption to an operating circuit. (*Hint:* Use parallel rather than series connections.)

1.58 A current amplifier supplies 1 mA to a load resistance of 1 kΩ. When the load resistance is increased to 12 kΩ, the output current decreases to 0.5 mA. What are the values of the short-circuit output current and the output resistance of the amplifier?

1.59 A transresistance amplifier for which $R_i = 100\ \Omega$, $R_o = 10\ \Omega$, and $R_m = 5\ \text{k}\Omega$ is to be connected between a 100-mV source with a resistance of 1 kΩ and a load of 1 kΩ. What are the values of current gain i_o/i_i, of voltage gain v_o/v_s, and of power gain expressed directly and in decibels?

1.60 A transconductance amplifier with $R_i = 2\ \text{k}\Omega$, $G_m = 20$ mA/V, and $R_o = 5\ \text{k}\Omega$ is fed with a voltage source having a source resistance of 500 Ω and is loaded with a 1-kΩ resistance. Find the voltage gain realized.

D **1.61 A designer is required to provide, across a 10-kΩ load, the weighted sum, $v_O = 10v_1 + 20v_2$, of input signals v_1 and v_2, each having a source resistance of 10 kΩ. She has a number of transconductance amplifiers for which the input and output resistances are both 10 kΩ and $G_m = 20$ mA/V, together with a selection of suitable resistors. Sketch an appropriate amplifier topology with additional resistors selected to provide the desired result. Your design should use the minimum number of amplifiers and resistors. (*Hint*: In your design, arrange to add currents.)

1.62 Figure P1.62 shows a transconductance amplifier whose output is *fed back* to its input.

(a) Find the input resistance R_{in} of the resulting one-port network. (*Hint*: Apply a test voltage v_x between the two input terminals, and find the current i_x drawn from the source. Then, $R_{in} \equiv v_x/i_x$.)

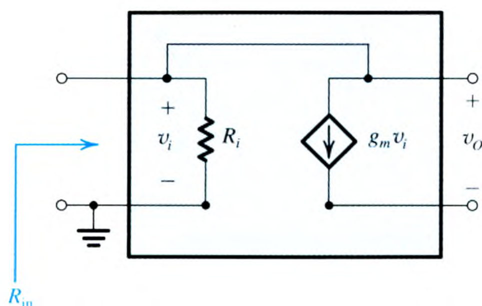


Figure P1.62

(b) Show that when driven by a voltage source, v_s, having source resistance $R_s = R_{in}$, the voltage gain $v_o/v_s = 0.5 \cdot v_o/v_i$ (This is another method for finding the input resistance.)

D 1.63 You are required to design an amplifier to sense the open-circuit output voltage of a transducer and to provide a proportional voltage across a load resistor. The equivalent source resistance of the transducer is specified to vary in the range of 1 kΩ to 10 kΩ. Also, the load resistance varies in the range of 1 kΩ to 10 kΩ. The change in load voltage corresponding to the specified change in R_s should be 10% at most. Similarly, the change in load voltage corresponding to the specified change in R_L should be limited to 10%. Also, corresponding to a 10-mV transducer open-circuit output voltage, the amplifier should provide a minimum of 1 V across the load. What type of amplifier is required? Sketch its circuit model, and specify the values of its parameters. Specify appropriate values for R_i and R_o of the form $1 \times 10^m\ \Omega$.

D 1.64 You are required to design an amplifier to sense the short-circuit output current of a light sensor and to provide a proportional voltage across a load resistor. The equivalent source resistance of the transducer is specified to vary in the range of 1 kΩ to 10 kΩ. Similarly, the load resistance is known to vary in the range of 1 kΩ to 10 kΩ. The change in load voltage corresponding to the specified change in R_s should be 10% at most. Similarly, the change in load voltage corresponding to the specified change in R_L is to be limited to 10%. Also, for a nominal transducer short-circuit output current of 10 μA, the amplifier is required to provide a minimum voltage across the load of 1 V. What type of amplifier is required? Sketch its circuit model, and specify the values of the model parameters. For R_i and R_o, specify appropriate values in the form $1 \times 10^m\ \Omega$.

1.65 For the circuit in Fig. P1.65, show that

$$\frac{v_c}{v_b} = \frac{-\beta R_L}{r_\pi + (\beta + 1)R_E}$$

and

$$\frac{v_e}{v_b} = \frac{R_E}{R_E + [r_\pi/(\beta + 1)]}$$

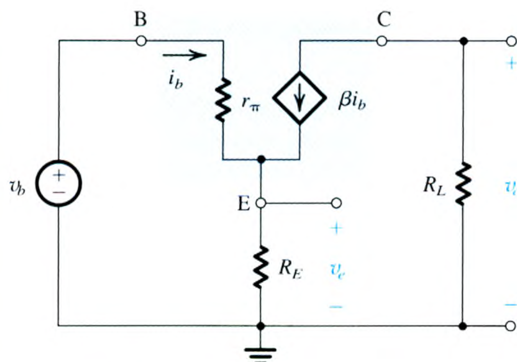


Figure P1.65

1.66 An amplifier with an input resistance of 5 kΩ, when driven by a current source of 1 μA and a source resistance of 200 kΩ, has a short-circuit output current of 5 mA. When the amplifier is used to drive a 2-kΩ load, an output voltage of 5 V is observed. When connected to a 1-kΩ load, give the values of the voltage gain, current gain, and power gain expressed as ratios and in decibels.

1.67 Figure P1.67(a) shows two transconductance amplifiers connected in a special configuration. Find v_o in terms of v_1 and v_2. Let $g_m = 100$ mA/V and $R = 5$ kΩ. If $v_1 = 1$ V, find the value of v_o. Also, find v_o for the case $v_1 = 1.01$ V and $v_2 = 0.99$ V. (*Note*: This circuit is called a **differential amplifier** and is given the symbol shown in Fig. P1.67(b). A particular type of differential amplifier known as an **operational amplifier** will be studied in Chapter 2.)

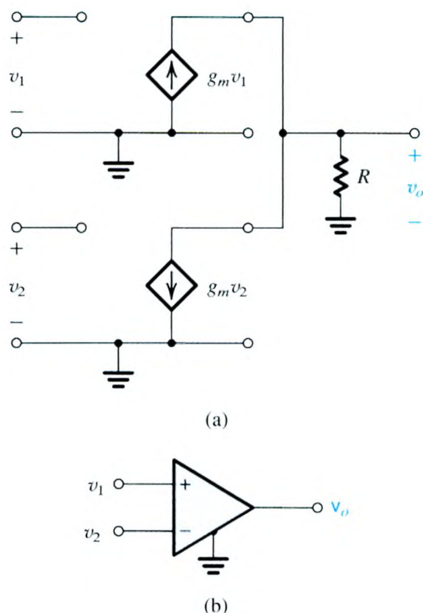


Figure P1.67

1.68 Any linear two-port network including linear amplifiers can be represented by one of four possible parameter sets, given in Appendix C. For the voltage amplifier, the most convenient representation is in terms of the g parameters. If the amplifier input port is labeled as port 1 and the output port as port 2, its g-parameter representation is described by the two equations:

$$I_1 = g_{11}V_1 + g_{12}I_2$$

$$V_2 = g_{21}V_1 + g_{22}I_2$$

Figure P1.68 shows an equivalent circuit representation of these two equations. By comparing this equivalent circuit to that of the voltage amplifier in Fig. 1.16(a), identify corresponding currents and voltages as well as the correspondence between the parameters of the amplifier equivalent circuit and the g parameters. Hence give the g parameter that corresponds to each of R_i, A_{vo}, and R_o. Notice that there is an additional g parameter with no correspondence in the amplifier equivalent circuit. Which one? What does it signify? What assumption did we make about the amplifier that resulted in the absence of this particular g parameter from the equivalent circuit in Fig. 1.16(a)?

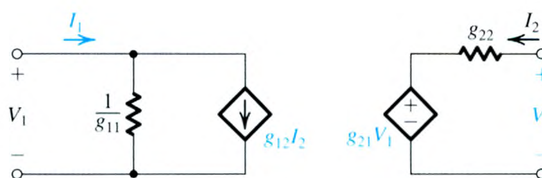


Figure P1.68

Section 1.6: Frequency Response of Amplifiers

1.69 Use the voltage-divider rule to derive the transfer functions $T(s) \equiv V_o(s)/V_i(s)$ of the circuits shown in Fig. 1.22,

and show that the transfer functions are of the form given at the top of Table 1.2.



1.70 Figure VE 1.4 shows a signal source connected to the input of an amplifier. Here R_s is the source resistance, and R_i and C_i are the input resistance and input capacitance, respectively, of the amplifier. Derive an expression for $V_i(s)/V_s(s)$, and show that it is of the low-pass STC type. Find the 3-dB frequency and dc gain for the case $R_s = 10\ \text{k}\Omega$, $R_i = 40\ \text{k}\Omega$, and $C_i = 5\ \text{pF}$.

1.71 For the circuit shown in Fig. P1.71, find the transfer function $T(s) = V_o(s)/V_i(s)$, and arrange it in the appropriate standard form from Table 1.2. Is this a high-pass or a low-pass network? What is its transmission at very high frequencies? [Estimate this directly, as well as by letting $s \to \infty$ in your expression for $T(s)$.] What is the corner frequency ω_0? For $R_1 = 20\ \text{k}\Omega$, $R_2 = 100\ \text{k}\Omega$, and $C = 0.1\ \mu\text{F}$, find f_0. What is the value of $|T(j\omega_0)|$?

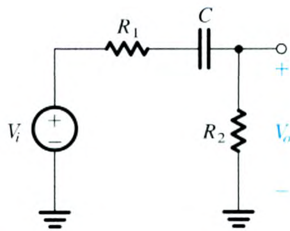


Figure P1.71

D 1.72 It is required to couple a voltage source V_s with a resistance R_s to a load R_L via a capacitor C. Derive an expression for the transfer function from source to load (i.e., V_l/V_s), and show that it is of the high-pass STC type. For $R_s = 4\ \text{k}\Omega$ and $R_L = 10\ \text{k}\Omega$, find the smallest coupling capacitor that will result in a 3-dB frequency no greater than 200 Hz.

1.73 Measurement of the frequency response of an amplifier yields the data in the following table:

f **(Hz)**	**|*T*| (dB)**	**∠*T*(°)**
0	60	0
100	60	
1000		−45
10^4	40	
10^5	20	
	0	

Provide plausible approximate values for the missing entries. Also, sketch and clearly label the magnitude frequency response (i.e., provide a Bode plot) for this amplifier.

1.74 Measurement of the frequency response of an amplifier yields the data in the following table:

f **(Hz)**		1	10^2	10^3	10^4	10^5	10^6	10^7	
|*T*| (dB)	0	40	77	80		77	60	40	0

Provide approximate plausible values for the missing table entries. Also, sketch and clearly label the magnitude frequency response (Bode plot) of this amplifier.

1.75 The unity-gain voltage amplifiers in Fig. P1.75 have infinite input resistances and zero output resistances and thus function as perfect buffers. Assuming that their gain is frequency independent, convince yourself that the overall gain V_o/V_i will drop by 3 dB below the value at dc at the frequency for which the gain of each *RC* circuit is 1.0 dB down. What is that frequency in terms of *CR*?

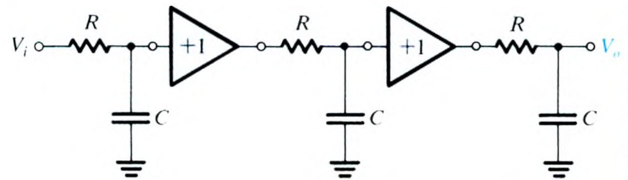


Figure P1.75

1.76 When a high-frequency transconductance amplifier whose output resistance is 100 kΩ is connected to a load capacitor, the measured 3-dB bandwidth of the amplifier is reduced from 5 MHz to 100 kHz. Estimate the value of the load capacitor. If the original cutoff frequency can be attributed to a small parasitic capacitor at the output node (i.e., between the output and ground), what would you estimate it to be?

D *1.77 A designer wishing to lower the overall upper 3-dB frequency of a three-stage amplifier to 5 kHz considers shunting one of two nodes to ground with a capacitor: Node A, at the output of the first stage, or Node B, at the output of the second stage. While measuring the overall frequency response of the amplifier, she connects a capacitor of 1 nF, first to node A and then to node B, lowering the 3-dB frequency from 3 MHz to 200 kHz and 40 kHz, respectively.

If she knows that each amplifier stage has an input resistance of 100 kΩ, what output resistance must the driving stage have at node A? At node B? What capacitor value should she connect to which node to solve her design problem most economically?

D 1.78 An amplifier with an input resistance of 100 kΩ and an output resistance of 1 kΩ is to be capacitively coupled to a 10-kΩ source and a 1-kΩ load. Available capacitors have values only of the form 1×10^{-n} F. What are the values of the smallest capacitors needed to ensure that the corner frequency associated with each is less than 100 Hz? What actual corner frequencies result? For the situation in which the basic amplifier has an open-circuit voltage gain (A_{vo}) of 100 V/V, find an expression for $T(s) = V_o(s)/V_s(s)$.

***1.79** A voltage amplifier has the transfer function

$$A_v = \frac{1000}{\left(1 + j\frac{f}{10^5}\right)\left(1 + \frac{10^2}{jf}\right)}$$

Using the Bode plots for low-pass and high-pass STC networks (Figs. 1.23 and 1.24), sketch a Bode plot for $|A_v|$. Give approximate values for the gain magnitude at $f = 10$ Hz, 10^2 Hz, 10^3 Hz, 10^4 Hz, 10^5 Hz, 10^6 Hz, 10^7 Hz, and 10^8 Hz. Find the bandwidth of the amplifier (defined as the frequency range over which the gain remains within 3 dB of the maximum value).

***1.80** For the circuit shown in Fig. P1.80, first evaluate $T_i(s) = V_i(s)/V_s(s)$ and the corresponding cutoff (corner) frequency. Second, evaluate $T_o(s) = V_o(s)/V_i(s)$ and the corresponding cutoff frequency. Put each of the transfer functions in the standard form (see Table 1.2), and combine them to form the overall transfer function, $T(s) = T_i(s) \times T_o(s)$. Provide a Bode plot for $|T(j\omega)|$. What is the bandwidth between 3-dB cutoff points?

D **1.81 A transconductance amplifier having the equivalent circuit shown in Table 1.1 is fed with a voltage source V_s having a source resistance R_s, and its output is connected to a load consisting of a resistance R_L in parallel with a capacitance C_L. For given values of R_s, R_L, and C_L, it is required to specify the values of the amplifier parameters R_i, G_m, and R_o to meet the following design constraints:

(a) At most, $x\%$ of the input signal is lost in coupling the signal source to the amplifier (i.e., $V_i \geq [1 - (x/100)]V_s$).
(b) The 3-dB frequency of the amplifier is equal to or greater than a specified value $f_{3\text{dB}}$.
(c) The dc gain V_o/V_s is equal to or greater than a specified value A_0.

Show that these constraints can be met by selecting

$$R_i \geq \left(\frac{100}{x} - 1\right)R_s$$

$$R_o \leq \frac{1}{2\pi f_{3\text{dB}} C_L - (1/R_L)}$$

$$G_m \geq \frac{A_0/[1 - (x/100)]}{(R_L \parallel R_o)}$$

Find R_i, R_o, and G_m for $R_s = 10$ kΩ, $x = 10\%$, $A_0 = 100$ V/V, $R_L = 10$ kΩ, $C_L = 20$ pF, and $f_{3\text{dB}} = 2$ MHz.

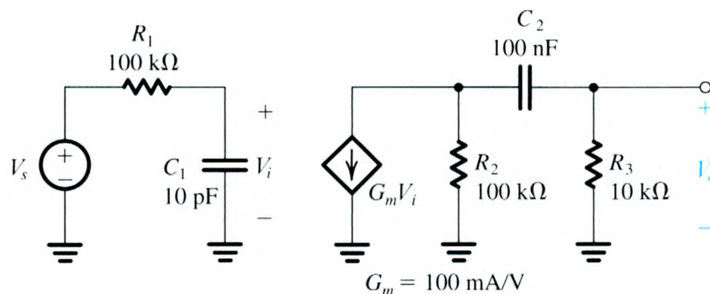


Figure P1.80

***1.82** Consider the circuit in Fig P1.82. It is desired to make the transfer function independent of frequency. Show this is achieved by selecting C_1 with a value $C_2(R_2/R_1)$. Under this condition the circuit is called a **compensated attenuator** and is frequently employed in the design of oscilloscope probes. Find the transmission of the compensated attenuator in terms of R_1 and R_2.

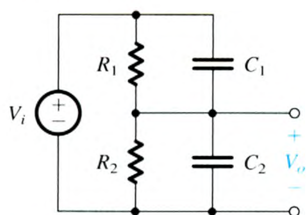


Figure P1.82

***1.83** An amplifier with a frequency response of the type shown in Fig. 1.21 is specified to have a phase shift of magnitude no greater than 5.7° over the amplifier bandwidth, which extends from 100 Hz to 1 kHz. You learn that the gain falloff at the low-frequency end is determined by the response of a high-pass STC circuit and that at the high-frequency end it is determined by a low-pass STC circuit. What do you expect the time constants of these two STC circuits to be? What is the drop in gain in decibels (relative to the maximum gain) at the two frequencies that define the amplifier bandwidth? What are the frequencies at which the drop in gain is 3 dB? (*Hint*: Refer to Figs. 1.23 and 1.24.)

CHAPTER 2

Operational Amplifiers

IN THIS CHAPTER YOU WILL LEARN

- The terminal characteristics of the ideal op amp.
- How to analyze circuits containing op amps, resistors, and capacitors.
- How to use op amps to design amplifiers having precise characteristics.
- How to design more sophisticated op-amp circuits, including summing amplifiers, instrumentation amplifiers, and integrators.
- Important nonideal characteristics of op amps and how these limit the performance of basic op-amp circuits.

Introduction

Having learned basic amplifier concepts and terminology, we are now ready to undertake the study of a circuit building block of universal importance: the operational amplifier (op amp).

THE OP AMP REVOLUTION

Early op amps were constructed from discrete components (vacuum tubes and then transistors, and resistors), and their cost was prohibitively high (tens of dollars). In the mid-1960s the first integrated-circuit (IC) op amp was produced (the μA 709), with its and resistors all on the same silicon chip. Although the op amp's characteristics were poor (by today's standards) and its price was still quite high, engineers started using op amps in large quantities, causing their price to drop dramatically. They also demanded better-quality op amps. Semiconductor manufacturers responded, and within the span of a few years high-quality op amps became available at extremely low prices (tens of cents) from a large number of suppliers.

One of the reasons for the popularity of the op amp is its versatility. As we will see, one can do almost anything with op amps! Equally important is the fact that the integrated circuit (IC) op amp has characteristics that closely approach the assumed ideal. This implies that it is quite easy to design circuits using the IC op amp. Also, op-amp circuits work at performance levels that are quite close to those predicted theoretically. It is for this reason that we are studying op amps at this early stage. It is expected that by the end of this chapter you should be able to successfully design nontrivial circuits using op amps.

An IC op amp is made up of many (about 20) transistors together with resistors, and (usually) one capacitor connected in a rather complex circuit. Since we have not yet studied transistor circuits, we will not discuss the circuit inside the op amp in this chapter. Instead we will treat the op amp as a circuit building block and study its terminal characteristics and applications. This approach is quite satisfactory in many op-amp applications. Nevertheless, for the more difficult and demanding applications it is quite useful to know what is inside the op-amp package. This topic will be studied in Chapter 13. More advanced applications of op amps will appear in later chapters.

2.1 The Ideal Op Amp

2.1.1 The Op-Amp Terminals

From a signal point of view the op amp has three terminals: two input terminals and one output terminal. Figure 2.1 shows the symbol we shall use to represent the op amp. Terminals 1 and 2 are input terminals, and terminal 3 is the output terminal. As explained in Section 1.4, amplifiers require dc power to operate. Most IC op amps require two dc power supplies, as shown in Fig. 2.2. Two terminals, 4 and 5, are brought out of the op-amp package and connected to a positive supply voltage V_{CC} and a negative supply voltage $-V_{EE}$, respectively. In Fig. 2.2(b) we explicitly show the two dc power supplies as batteries with a common ground. In some op amps, the negative supply terminal may be connected directly to ground ($V_{EE} = 0$). In this book, the common

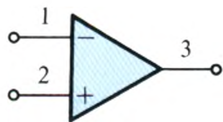


Figure 2.1 Circuit symbol for the op amp.

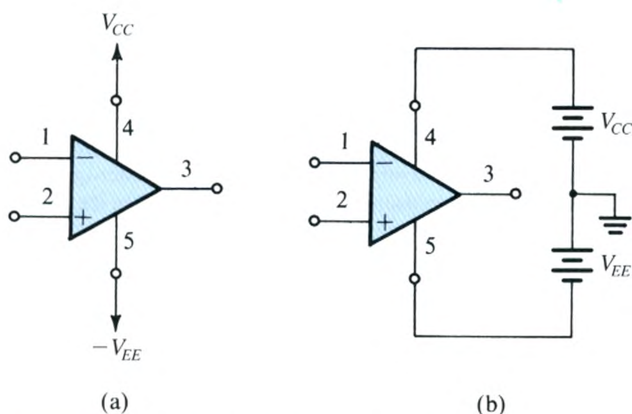


Figure 2.2 The op amp shown connected to dc power supplies.

terminals between the two power supplies is considered to be the circuit ground (see Fig. 2.2b). In what follows we will not, for simplicity, explicitly show the op-amp power supplies.

In addition to the three signal terminals and the two power-supply terminals, an op amp may have other terminals for specific purposes. These other terminals can include terminals for frequency compensation and terminals for offset nulling; both functions will be explained in later sections.

EXERCISE

2.1 What is the minimum number of terminals required by a single op amp? What is the minimum number of terminals required on an integrated-circuit package containing four op amps (called a quad op amp)?
Ans. 5; 14

2.1.2 Function and Characteristics of the Ideal Op Amp

The op amp is designed to sense the difference between the voltage signals applied at its two input terminals (i.e., the quantity $v_2 - v_1$), multiply this by a number A, and cause the resulting voltage $A(v_2 - v_1)$ to appear at output terminal 3. Thus $v_3 = A(v_2 - v_1)$. Here we

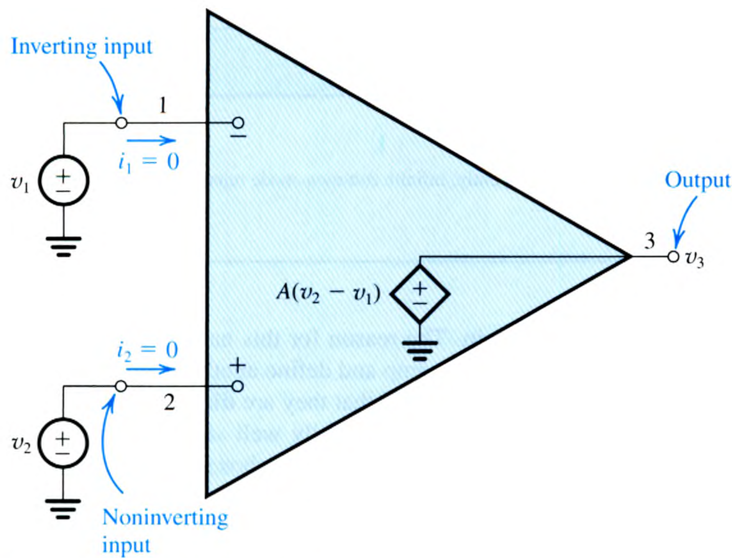


Figure 2.3 Equivalent circuit of the ideal op amp.

should emphasize that when we talk about the voltage at a terminal we mean the voltage between that terminal and ground; thus v_1 means the voltage applied between terminal 1 and ground.

The ideal op amp is not supposed to draw any input current; that is, the signal current into terminal 1 and the signal current into terminal 2 are both zero. In other words, *the input impedance of an ideal op amp is supposed to be infinite.* How about the output terminal 3? This terminal is supposed to act as the output terminal of an ideal voltage source. That is, the voltage between terminal 3 and ground will always be equal to $A(v_2 - v_1)$, independent of the current that may be drawn from terminal 3 into a load impedance. In other words, *the output impedance of an ideal op amp is supposed to be zero.*

Putting together all of the above, we arrive at the equivalent circuit model shown in Fig. 2.3. Note that the output is in phase with (has the same sign as) v_2 and is out of phase with (has the opposite sign of) v_1. For this reason, input terminal 1 is called the **inverting input terminal** and is distinguished by a "−" sign, while input terminal 2 is called the **noninverting input terminal** and is distinguished by a "+" sign.

As the above description makes clear, the op amp responds only to the *difference* signal $v_2 - v_1$ and hence ignores any signal *common* to both inputs. That is, if $v_1 = v_2 = 1$ V, then the output will (ideally) be zero. We call this property **common-mode rejection**, and we conclude that an ideal op amp has zero common-mode gain or, equivalently, infinite common-mode rejection. We will have more to say about this point later. For the time being note that the op amp is a **differential-input, single-ended-output** amplifier, with the latter term referring to the fact that the output appears between terminal 3 and ground.[1] Gain A is called the **differential gain**, for obvious reasons. Perhaps not so obvious is another

[1]Some op amps are designed to have differential outputs. This topic will not be discussed in this book. Rather, we confine ourselves here to single-ended-output op amps, which constitute the vast majority of commercially available op amps.

Table 2.1 Characteristics of the Ideal Op Amp

1. Infinite input impedance
2. Zero output impedance
3. Zero common-mode gain or, equivalently, infinite common-mode rejection
4. Infinite open-loop gain *A*
5. Infinite bandwidth

name we give it: the **open-loop gain**. The reason for this name will become obvious later, when we "close the loop" around the op amp and define another gain, the closed-loop gain.

An important characteristic of op amps is that they are **direct-coupled** or **dc amplifiers**, where dc stands for direct-coupled (it could equally well stand for direct current, since a direct-coupled amplifier is one that amplifies signals whose frequency is as low as zero). The fact that op amps are direct-coupled devices will allow us to use them in many important applications. For example, they may be used to amplify signals that change very slowly, such as from a temperature sensor.

How about bandwidth? The ideal op amp has a gain A that remains constant down to zero frequency and up to infinite frequency. That is, ideal op amps will amplify signals of any frequency with equal gain, and are thus said to have *infinite bandwidth.*

We have discussed all of the properties of the ideal op amp except for one, which in fact is the most important. This has to do with the value of A. *The ideal op amp should have a gain A whose value is very large and ideally infinite.* You may justifiably ask: If the gain A is infinite, how are we going to use the op amp? The answer is very simple: In almost all applications the op amp will *not* be used alone in a so-called open-loop configuration. Rather, we will use other components to apply feedback to close the loop around the op amp, as will be illustrated in detail in Section 2.2.

For future reference, Table 2.1 lists the characteristics of the ideal op amp.

2.1.3 Differential and Common-Mode Signals

The differential input signal v_{Id} is simply the difference between the two input signals v_1 and v_2; that is,

$$v_{Id} = v_2 - v_1 \tag{2.1}$$

The common-mode input signal v_{Icm} is the average of the two input signals v_1 and v_2; namely,

$$v_{Icm} = \tfrac{1}{2}(v_1 + v_2) \tag{2.2}$$

Equations (2.1) and (2.2) can be used to express the input signals v_1 and v_2 in terms of their differential and common-mode components as follows:

$$v_1 = v_{Icm} - v_{Id}/2 \tag{2.3}$$

and

$$v_2 = v_{Icm} + v_{Id}/2 \tag{2.4}$$

These equations can in turn lead to the pictorial representation in Fig. 2.4.

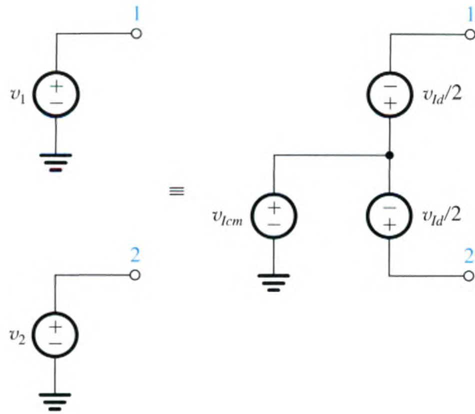


Figure 2.4 Representation of the signal sources v_1 and v_2 in terms of their differential and common-mode components.

EXERCISES

2.2 Consider an op amp that is ideal except that its open-loop gain $A = 10^3$. The op amp is used in a feedback circuit, and the voltages appearing at two of its three signal terminals are measured. In each of the following cases, use the measured values to find the expected value of the voltage at the third terminal. Also give the differential and common-mode input signals in each case. (a) $v_2 = 0$ V and $v_3 = 4$ V; (b) $v_2 = +2$ V and $v_3 = -10$ V; (c) $v_1 = 2.002$ V and $v_2 = 1.998$ V; (d) $v_1 = -1.2$ V and $v_3 = -1.2$ V.

Ans. (a) $v_1 = -4$ mV, $v_{Id} = 4$ mV, $v_{Icm} = -2$ mV; (b) $v_1 = +2.01$ V, $v_{Id} = -10$ mV, $v_{Icm} = 2.005$ V $\simeq$ 2 V; (c) $v_3 = -4$ V, $v_{Id} = -4$ mV, $v_{Icm} = 2$ V; (d) $v_2 = -1.2012$ V, $v_{Id} = -1.2$ mV, $v_{Icm} \simeq -1.2$ V

2.3 The internal circuit of a particular op amp can be modeled by the circuit shown in Fig. E2.3. Express v_3 as a function of v_1 and v_2. For the case $G_m = 20$ mA/V, $R = 5$ kΩ, and $\mu = 50$, find the value of the open-loop gain A.

Ans. $v_3 = \mu G_m R(v_2 - v_1)$; $A = 5{,}000$ V/V or 74 dB

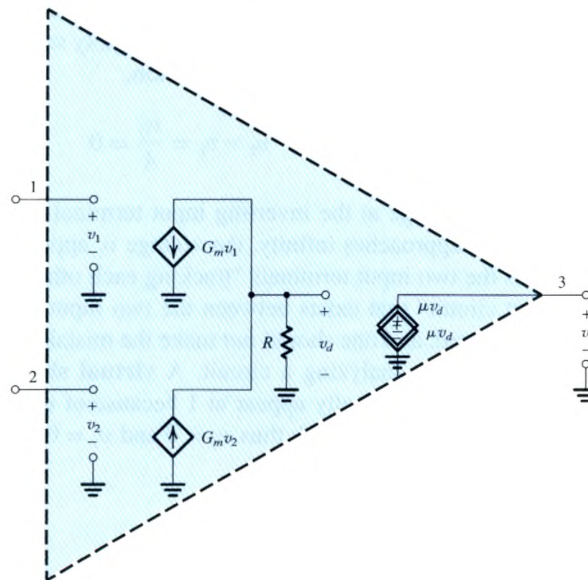


Figure E2.3

2.2 The Inverting Configuration

As mentioned above, op amps are not used alone; rather, the op amp is connected to passive components in a feedback circuit. There are two such basic circuit configurations employing an op amp and two resistors: the inverting configuration, which is studied in this section, and the noninverting configuration, which we shall study in the next section.

Figure 2.5 shows the inverting configuration. It consists of one op amp and two resistors, R_1 and R_2. Resistor R_2 is connected from the output terminal of the op amp, terminal 3, *back* to the *inverting* or *negative* input terminal, terminal 1. We speak of R_2 as applying **negative feedback**; if R_2 were connected between terminals 3 and 2 we would have called this **positive feedback**, which generally causes the op-amp output to saturate rather than providing a stable gain. Therefore, the inverting configuration will always be connected with negative feedback. Note also that R_2 *closes the loop* around the op amp. In addition to adding R_2, we have grounded terminal 2 and connected a resistor R_1 between terminal 1 and an input signal source with a voltage v_I. The output of the overall circuit is taken at terminal 3 (i.e., between terminal 3 and ground). Terminal 3 is, of course, a convenient point from which to take the output, since the impedance level there is ideally zero. Thus the voltage v_O will not depend on the value of the current that might be supplied to a load impedance connected between terminal 3 and ground.

2.2.1 The Closed-Loop Gain

We now wish to analyze the circuit in Fig. 2.5 to determine the **closed-loop gain** G, defined as

$$G \equiv \frac{v_O}{v_I}$$

We will do so assuming the op amp to be ideal. Figure 2.6(a) shows the equivalent circuit, and the analysis proceeds as follows: The gain A is very large (ideally infinite). If we assume that the circuit is "working" and producing a finite output voltage at terminal 3, then the voltage between the op-amp input terminals should be negligibly small and ideally zero. Specifically, if we call the output voltage v_O, then, by definition,

$$v_2 - v_1 = \frac{v_O}{A} = 0$$

It follows that the voltage at the inverting input terminal (v_1) is given by $v_1 = v_2$. That is, because the gain A approaches infinity, the voltage v_1 approaches and ideally equals v_2. We speak of this as the two input terminals "tracking each other in potential." We also speak of a "virtual short circuit" that exists between the two input terminals. Here the word *virtual* should be emphasized, and one should *not* make the mistake of physically shorting terminals 1 and 2 together while analyzing a circuit. A **virtual short circuit** means that whatever voltage is at 2 will automatically appear at 1 because of the infinite gain A. But terminal 2 happens to be connected to ground; thus $v_2 = 0$ and $v_1 = 0$. We speak of terminal 1 as being

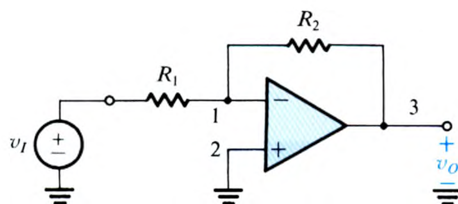
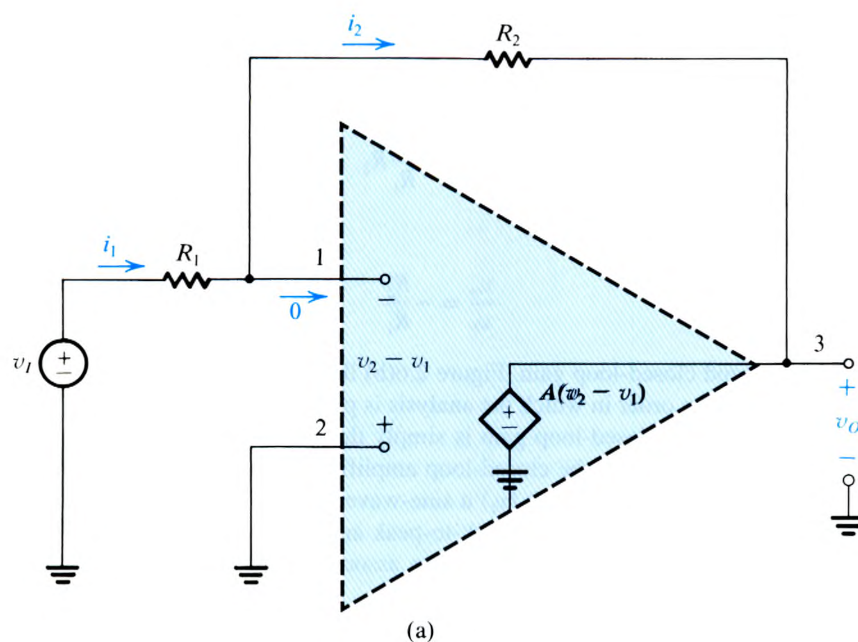


Figure 2.5 The inverting closed-loop configuration.



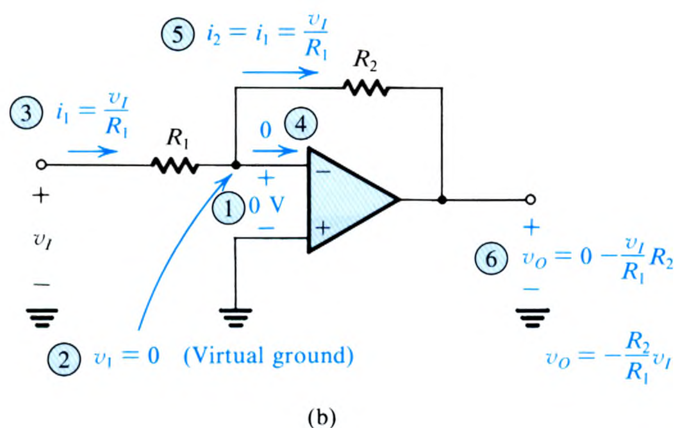


Figure 2.6 Analysis of the inverting configuration. The circled numbers indicate the order of the analysis steps.

a **virtual ground**—that is, having zero voltage but not physically connected to ground.

Now that we have determined v_1 we are in a position to apply Ohm's law and find the current i_1 through R_1 (see Fig. 2.6) as follows:

$$i_1 = \frac{v_I - v_1}{R_1} = \frac{v_I - 0}{R_1} = \frac{v_I}{R_1}$$

Where will this current go? It cannot go into the op amp, since the ideal op amp has an infinite input impedance and hence draws zero current. It follows that i_1 will have to flow through

R_2 to the low-impedance terminal 3. We can then apply Ohm's law to R_2 and determine v_O; that is,

$$v_O = v_1 - i_1 R_2$$
$$= 0 - \frac{v_I}{R_1} R_2$$

Thus,

$$\frac{v_O}{v_I} = -\frac{R_2}{R_1}$$

which is the required closed-loop gain. Figure 2.6(b) illustrates these steps and indicates by the circled numbers the order in which the analysis is performed.

We thus see that the closed-loop gain is simply the ratio of the two resistances R_2 and R_1. The minus sign means that the closed-loop amplifier provides signal inversion. Thus if $R_2/R_1 = 10$ and we apply at the input (v_I) a sine-wave signal of 1 V peak-to-peak, then the output v_O will be a sine wave of 10 V peak-to-peak and phase-shifted 180° with respect to the input sine wave. Because of the minus sign associated with the closed-loop gain, this configuration is called the **inverting configuration**.

The fact that the closed-loop gain depends entirely on external passive components (resistors R_1 and R_2) is very significant. It means that we can make the closed-loop gain as accurate as we want by selecting passive components of appropriate accuracy. It also means that the closed-loop gain is (ideally) independent of the op-amp gain. This is a dramatic illustration of negative feedback: We started out with an amplifier having very large gain A, and by applying negative feedback we have obtained a closed-loop gain R_2/R_1 that is much smaller than A but is stable and predictable. That is, we are trading gain for accuracy.

Video Example VE 2.1

An ideal op amp is connected as shown in Fig. 2.5 with $R_1 = 10\ \text{k}\Omega$ and $R_2 = 100\ \text{k}\Omega$. A symmetrical square wave with levels of 0 V and 1 V is applied at the input. Sketch and clearly label the waveform of the resulting output voltage. What is its average value? What is its highest value? What is its lowest value? Also sketch the current waveform through R_1, labeling its maximum, minimum, and average.

VE 2.1

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 2.16

2.2.2 Effect of Finite Open-Loop Gain

The points just made are more clearly illustrated by deriving an expression for the closed-loop gain under the assumption that the op-amp open-loop gain A is finite. Figure 2.7 shows the analysis. If we denote the output voltage v_O, then the voltage between the two input terminals

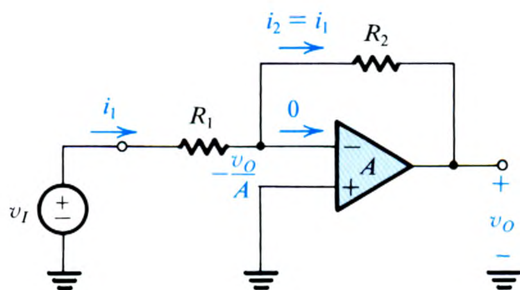


Figure 2.7 Analysis of the inverting configuration taking into account the finite open-loop gain of the op amp.

of the op amp will be v_O/A. Since the positive input terminal is grounded, the voltage at the negative input terminal must be $-v_O/A$. The current i_1 through R_1 can now be found from

$$i_1 = \frac{v_I - (-v_O/A)}{R_1} = \frac{v_I + v_O/A}{R_1}$$

The infinite input impedance of the op amp forces the current i_1 to flow entirely through R_2. The output voltage v_O can thus be determined from

$$\begin{aligned} v_O &= -\frac{v_O}{A} - i_1 R_2 \\ &= -\frac{v_O}{A} - \left(\frac{v_I + v_O/A}{R_1}\right) R_2 \end{aligned}$$

Collecting terms, the closed-loop gain G is found as

$$G \equiv \frac{v_O}{v_I} = \frac{-R_2/R_1}{1 + (1 + R_2/R_1)/A} \tag{2.5}$$

We note that as A approaches ∞, G approaches the ideal value of $-R_2/R_1$. Also, from Fig. 2.7 we see that as A approaches ∞, the voltage at the inverting input terminal approaches zero. This is the virtual-ground assumption we used in our earlier analysis when the op amp was assumed to be ideal. Finally, note that Eq. (2.5) in fact indicates that to minimize the dependence of the closed-loop gain G on the value of the open-loop gain A, we should make

$$1 + \frac{R_2}{R_1} \ll A$$

If A is not much greater than $(1 + R_2/R_1)$, the closed-loop gain G will decrease below its ideal value. In practice, A (and hence G) are frequency dependent so that the ideal behavior is approximated only up to some maximum frequency.

Example 2.1

Consider the inverting configuration with $R_1 = 1\ \text{k}\Omega$ and $R_2 = 100\ \text{k}\Omega$, that is, having an ideal closed-loop gain of -100.

(a) Find the closed-loop gain for the cases $A = 10^3, 10^4$, and 10^5. In each case determine the percentage error in the magnitude of G relative to the ideal value of R_2/R_1 (obtained with $A = \infty$). Also determine the voltage v_1 that appears at the inverting input terminal when $v_I = 0.1$ V.

(b) If the open-loop gain A changes from 100,000 to 50,000 (i.e., drops by 50%), what is the corresponding percentage change in the magnitude of the closed-loop gain G?

Solution

(a) Substituting the given values in Eq. (2.5), we obtain the values given in the following table, where the percentage error ϵ is defined as

$$\epsilon \equiv \frac{|G| - (R_2/R_1)}{(R_2/R_1)} \times 100$$

The values of v_1 are obtained from $v_1 = -v_O/A = Gv_I/A$ with $v_I = -0.1$ V.

A	$\lvert G\rvert$	ϵ	v_1
10^3	90.83	−9.17%	−9.08 mV
10^4	99.00	−1.00%	−0.99 mV
10^5	99.90	−0.10%	−0.10 mV

(b) Using Eq. (2.5), we find that for $A = 50{,}000, |G| = 99.80$. Thus a -50% change in the open-loop gain results in a change in $|G|$ from 99.90 to 99.80, which is only -0.1%!

2.2.3 Input and Output Resistances

Assuming an ideal op amp with infinite open-loop gain, the input resistance of the closed-loop inverting amplifier of Fig. 2.5 is simply equal to R_1. This can be seen from Fig. 2.6(b), where

$$R_i \equiv \frac{v_I}{i_1} = \frac{v_I}{v_I/R_1} = R_1$$

Now recall that in Section 1.5 we learned that the amplifier input resistance forms a voltage divider with the resistance of the source that feeds the amplifier. Thus, to avoid the loss of signal strength, voltage amplifiers are required to have high input resistance. In the case of the inverting op-amp configuration we are studying, to make R_i high we should select a high value for R_1. However, if the required gain R_2/R_1 is also high, then R_2 could become impractically large (e.g., greater than a few megohms). We may conclude that the inverting configuration suffers from a low input resistance. A solution to this problem is discussed in Example 2.2 below.

Since the output of the inverting configuration is taken at the terminals of the ideal voltage source $A(v_2 - v_1)$ (see Fig. 2.6a), it follows that the output resistance of the closed-loop amplifier is zero.

Example 2.2

Assuming the op amp to be ideal, derive an expression for the closed-loop gain v_O/v_I of the circuit shown in Fig. 2.8. Use this circuit to design an inverting amplifier with a gain of 100 and an input resistance of 1 MΩ. Assume that for practical reasons it is required not to use resistors greater than 1 MΩ. Compare your design with that based on the inverting configuration of Fig. 2.5.

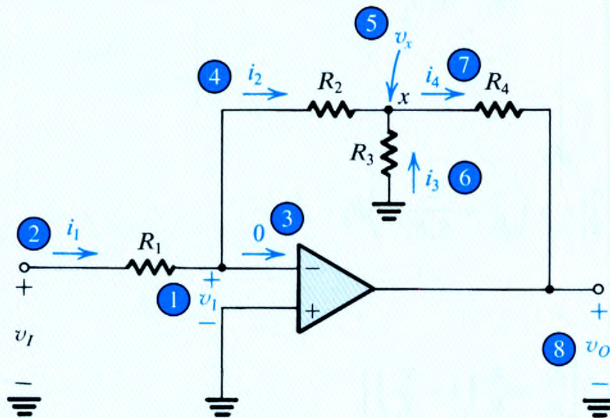


Figure 2.8 Circuit for Example 2.2. The circled numbers indicate the sequence of the steps in the analysis.

Solution

The analysis begins at the inverting input terminal of the op amp, where the voltage is

$$v_1 = \frac{-v_O}{A} = \frac{-v_O}{\infty} = 0$$

Here we have assumed that the circuit is "working" and producing a finite output voltage v_O. Knowing v_1, we can determine the current i_1 as follows:

$$i_1 = \frac{v_I - v_1}{R_1} = \frac{v_I - 0}{R_1} = \frac{v_I}{R_1}$$

Since zero current flows into the inverting input terminal, all of i_1 will flow through R_2, and thus

$$i_2 = i_1 = \frac{v_I}{R_1}$$

Now we can determine the voltage at node x:

$$v_x = v_1 - i_2 R_2 = 0 - \frac{v_I}{R_1} R_2 = -\frac{R_2}{R_1} v_I$$

Example 2.2 *continued*

This in turn enables us to find the current i_3:

$$i_3 = \frac{0 - v_x}{R_3} = \frac{R_2}{R_1 R_3} v_I$$

Next, a node equation at x yields i_4:

$$i_4 = i_2 + i_3 = \frac{v_I}{R_1} + \frac{R_2}{R_1 R_3} v_I$$

Finally, we can determine v_O from

$$\begin{aligned} v_O &= v_x - i_4 R_4 \\ &= -\frac{R_2}{R_1} v_I - \left(\frac{v_I}{R_1} + \frac{R_2}{R_1 R_3} v_I \right) R_4 \end{aligned}$$

Thus the voltage gain is given by

$$\frac{v_O}{v_I} = -\left[\frac{R_2}{R_1} + \frac{R_4}{R_1} \left(1 + \frac{R_2}{R_3} \right) \right]$$

which can be written in the form

$$\frac{v_O}{v_I} = -\frac{R_2}{R_1} \left(1 + \frac{R_4}{R_2} + \frac{R_4}{R_3} \right)$$

Now, since an input resistance of 1 MΩ is required, we select $R_1 = 1$ MΩ. Then, with the limitation of using resistors no greater than 1 MΩ, the maximum value possible for the first factor in the gain expression is 1 and is obtained by selecting $R_2 = 1$ MΩ. To obtain a gain of -100, R_3 and R_4 must be selected so that the second factor in the gain expression is 100. If we select the maximum allowed (in this example) value of 1 MΩ for R_4, then the required value of R_3 can be calculated to be 10.2 kΩ. Thus this circuit utilizes three 1-MΩ resistors and a 10.2-kΩ resistor. In comparison, if the inverting configuration were used with $R_1 = 1$ MΩ we would have required a feedback resistor of 100 MΩ, an impractically large value!

Before leaving this example it is insightful to inquire into the mechanism by which the circuit is able to realize a large voltage gain without using large resistances in the feedback path. Toward that end, observe that because of the virtual ground at the inverting input terminal of the op amp, R_2 and R_3 are in effect in parallel. Thus, by making R_3 lower than R_2 by, say, a factor k (i.e., where $k > 1$), R_3 is forced to carry a current k-times that in R_2. Thus, while $i_2 = i_1$, $i_3 = ki_1$ and $i_4 = (k+1)i_1$. It is the current multiplication by a factor of $(k+1)$ that enables a large voltage drop to develop across R_4 and hence a

large v_O without using a large value for R_4. Notice also that the current through R_4 is independent of the value of R_4. It follows that the circuit can be used as a current amplifier as shown in Fig. 2.9.

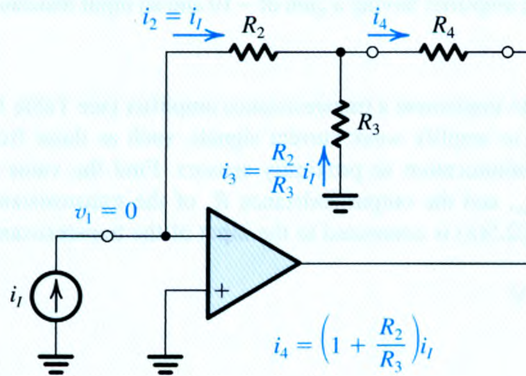


Figure 2.9 A current amplifier based on the circuit of Fig. 2.8. The amplifier delivers its output current to R_4. It has a current gain of $(1 + R_2/R_3)$, a zero input resistance, and an infinite output resistance. The load (R_4), however, must be floating (i.e., neither of its two terminals can be connected to ground).

Video Example VE 2.2

Assuming the op amp to be ideal, we must design the circuit in Fig. VE2.2 to implement a current amplifier with gain $i_L/i_I = 6$ A/A.

(a) Find the required value for R.
(b) What are the input and the output resistance of this current amplifier?
(c) If $R_L = 1$ kΩ and the op amp operates in an ideal manner as long as V_O is in the range ±5 V, what range of i_I is possible?
(d) If the amplifier is fed with a current source having a current of 0.2 mA and a source resistance of 10 kΩ, find i_L.

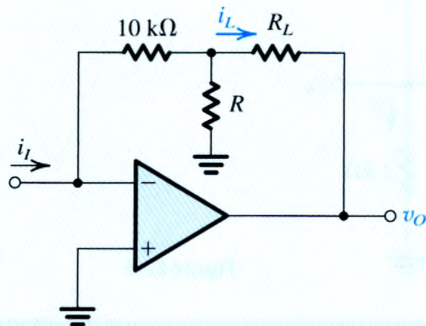


Figure VE2.2 Circuit for Video Example VE 2.2.

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

VE 2.2

Related end-of-chapter problem: 2.36

EXERCISES

D2.4 Use the circuit of Fig. 2.5 to design an inverting amplifier having a gain of −10 and an input resistance of 100 kΩ. Give the values of R_1 and R_2.
Ans. $R_1 = 100\ \text{k}\Omega$; $R_2 = 1\ \text{M}\Omega$

2.5 The circuit shown in Fig. E2.5(a) can be used to implement a transresistance amplifier (see Table 1.1 in Section 1.5). Such circuits are often used to amplify weak current signals, such as those from photodiodes that detect light for optical communication or proximity sensors. Find the value of the input resistance R_i, the transresistance R_m, and the output resistance R_o of the transresistance amplifier. If the signal source shown in Fig. E2.5(b) is connected to the input of the transresistance amplifier, find the amplifier output voltage.
Ans. $R_i = 0$; $R_m = -10\ \text{k}\Omega$; $R_o = 0$; $v_O = -5$ V

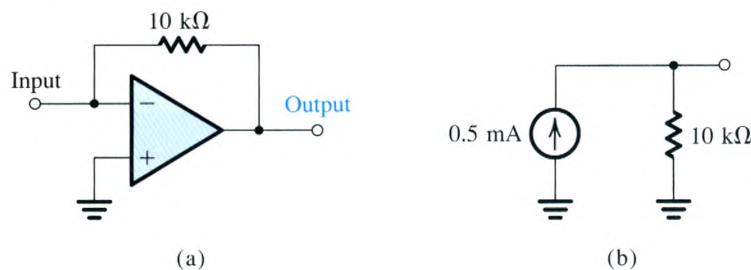


Figure E2.5

2.6 For the circuit in Fig. E2.6 determine the values of v_1, i_1, i_2, v_O, i_L, and i_O. Also determine the voltage gain v_O/v_I, current gain i_L/i_I, and power gain P_O/P_I.
Ans. 0 V; 2 mA; 2 mA; −10 V; −10 mA; −12 mA; −5 V/V (14 dB); −5 A/A (14 dB); 25 W/W (14 dB)

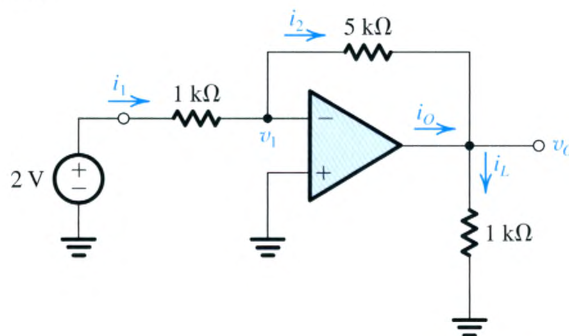


Figure E2.6

2.2.4 An Important Application: The Weighted Summer

A very important application of the inverting configuration is the weighted-summer circuit shown in Fig. 2.10. Here we have a resistance R_f in the negative-feedback path (as before),

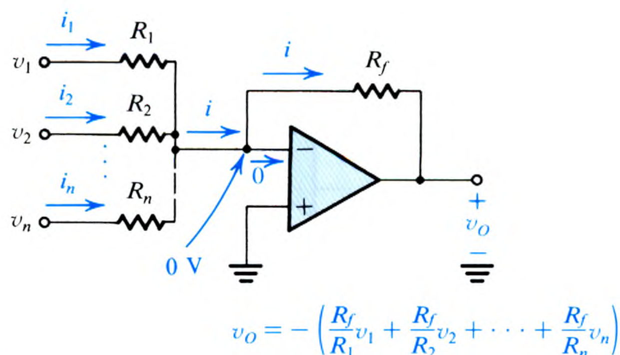


Figure 2.10 A weighted summer.

but we have a number of input signals $v_1, v_2, \ldots, v_n$ each applied to a corresponding resistor $R_1, R_2, \ldots, R_n$, which are connected to the inverting terminal of the op amp. From our previous discussion, the ideal op amp will have a virtual ground appearing at its negative input terminal. Ohm's law then tells us that the currents $i_1, i_2, \ldots, i_n$ are given by

$$i_1 = \frac{v_1}{R_1}, \quad i_2 = \frac{v_2}{R_2}, \quad \ldots, \quad i_n = \frac{v_n}{R_n}$$

All these currents sum together to produce the current i,

$$i = i_1 + i_2 + \cdots + i_n \tag{2.6}$$

which will be forced to flow through R_f (since no current flows into the input terminals of an ideal op amp). The output voltage v_O may now be determined by another application of Ohm's law,

$$v_O = 0 - iR_f = -iR_f$$

Thus,

$$v_O = -\left(\frac{R_f}{R_1}v_1 + \frac{R_f}{R_2}v_2 + \cdots + \frac{R_f}{R_n}v_n\right) \tag{2.7}$$

That is, the output voltage is a weighted sum of the input signals $v_1, v_2, \ldots, v_n$. This circuit is therefore called a **weighted summer**. Note that each summing coefficient may be independently adjusted by adjusting the corresponding "feed-in" resistor (R_1 to R_n). This nice property, which greatly simplifies circuit adjustment, is a direct consequence of the virtual ground that exists at the inverting op-amp terminal. As the reader will soon come to appreciate, virtual grounds are extremely "handy." In the weighted summer of Fig. 2.10 all the summing coefficients must be of the same sign. The need occasionally arises for summing signals with opposite signs. Such a function can be implemented, however, using two op amps as shown in Fig. 2.11. Assuming ideal op amps, it can be easily shown that the output voltage is given by

$$v_O = v_1\left(\frac{R_a}{R_1}\right)\left(\frac{R_c}{R_b}\right) + v_2\left(\frac{R_a}{R_2}\right)\left(\frac{R_c}{R_b}\right) - v_3\left(\frac{R_c}{R_3}\right) - v_4\left(\frac{R_c}{R_4}\right) \tag{2.8}$$

Weighted summers are utilized in a variety of applications including in the design of audio systems, where they can be used in mixing signals originating from different musical instruments.

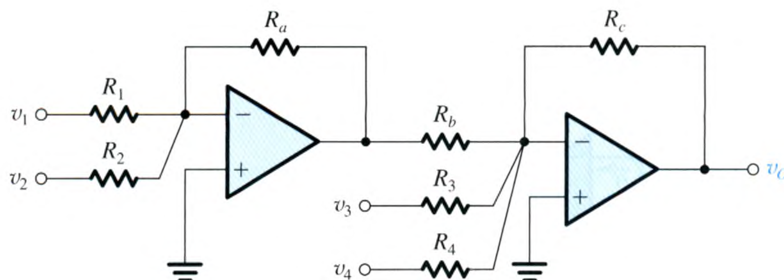


Figure 2.11 A weighted summer capable of implementing summing coefficients of both signs.

EXERCISES

D2.7 Design an inverting op-amp circuit to form the weighted sum v_O of two inputs v_1 and v_2. It is required that $v_O = -(v_1 + 4v_2)$. Choose values for R_1, R_2, and R_f so that for a maximum output voltage of 4 V the current in the feedback resistor will not exceed 1 mA.
Ans. A possible choice: $R_1 = 4\text{ k}\Omega, R_2 = 1\text{ k}\Omega$, and $R_f = 4\text{ k}\Omega$

D2.8 Use the idea presented in Fig. 2.11 to design a weighted summer that provides

$$v_O = 2v_1 + v_2 - 4v_3$$

Ans. A possible choice: $R_1 = 5\text{ k}\Omega$, $R_2 = 10\text{ k}\Omega$, $R_a = 10\text{ k}\Omega$, $R_b = 10\text{ k}\Omega$, $R_3 = 2.5\text{ k}\Omega$, $R_c = 10\text{ k}\Omega$

2.3 The Noninverting Configuration

The second closed-loop configuration we shall study is shown in Fig. 2.12. Here the input signal v_I is applied directly to the positive input terminal of the op amp while one terminal of R_1 is connected to ground.

2.3.1 The Closed-Loop Gain

Analysis of the noninverting circuit to determine its closed-loop gain (v_O/v_I) is illustrated in Fig. 2.13. Again the order of the steps in the analysis is indicated by circled numbers. Assuming that the op amp is ideal with infinite gain, a virtual short circuit exists between its two input terminals. Hence the difference input signal is

$$v_{Id} = \frac{v_O}{A} = 0 \quad \text{for } A = \infty$$

Thus the voltage at the inverting input terminal will be equal to that at the noninverting input terminal, which is the applied voltage v_I. The current through R_1 can then be determined as v_I/R_1. Because of the infinite input impedance of the op amp, this current will flow through

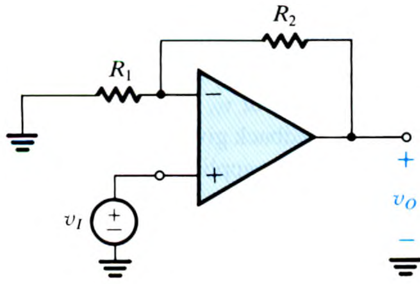


Figure 2.12 The noninverting configuration.

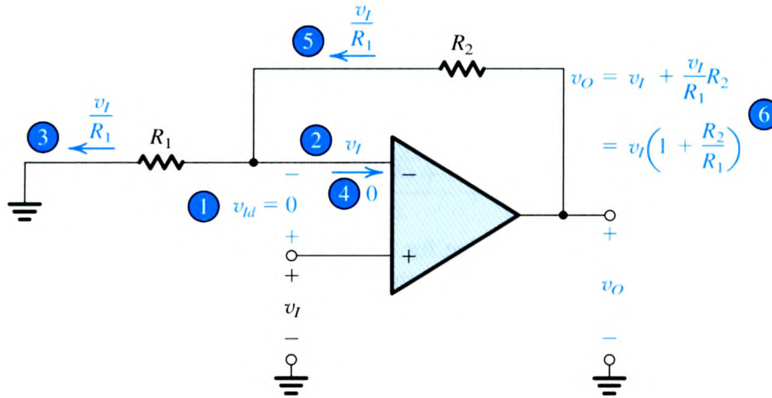


Figure 2.13 Analysis of the noninverting circuit. The sequence of the steps in the analysis is indicated by the circled numbers.

R_2, as shown in Fig. 2.13. Now the output voltage can be determined from

$$v_O = v_I + \left(\frac{v_I}{R_1}\right)R_2$$

which yields

$$\frac{v_O}{v_I} = 1 + \frac{R_2}{R_1} \tag{2.9}$$

To gain further insight into the operation of the noninverting configuration, consider this point: Since the current into the op-amp inverting input is zero, the circuit composed of R_1 and R_2 acts in effect as a voltage divider feeding a fraction of the output voltage back to the inverting input terminal of the op amp; that is,

$$v_1 = v_O\left(\frac{R_1}{R_1 + R_2}\right) \tag{2.10}$$

Then the infinite op-amp gain and the resulting virtual short circuit between the two input terminals of the op amp forces this voltage to be equal to that applied at the positive input terminal; thus,

$$v_O\left(\frac{R_1}{R_1 + R_2}\right) = v_I$$

which yields the gain expression given in Eq. (2.9).

At this point, let us reflect further on the action of the negative feedback present in the noninverting circuit of Fig. 2.12. Let v_I increase. Such a change in v_I will cause v_{Id} to increase,

and v_O will correspondingly increase as a result of the high (ideally infinite) gain of the op amp. However, a fraction of the increase in v_O will be fed back to the inverting input terminal of the op amp through the (R_1, R_2) voltage divider. The result of this feedback will be to counteract the increase in v_{Id}, driving v_{Id} back to zero, albeit at a higher value of v_O that corresponds to the increased value of v_I. This *degenerative* action of negative feedback gives it the alternative name **degenerative feedback**. Finally, note that the argument above applies equally well if v_I decreases. A formal and detailed study of feedback is presented in Chapter 11.

2.3.2 Effect of Finite Open-Loop Gain

As we have done for the inverting configuration, we now consider the effect of the finite op-amp open-loop gain A on the gain of the noninverting configuration. Assuming the op amp to be ideal except for having a finite open-loop gain A, it can be shown that the closed-loop gain of the noninverting amplifier circuit of Fig. 2.12 is given by

$$G \equiv \frac{v_O}{v_I} = \frac{1 + (R_2/R_1)}{1 + \dfrac{1 + (R_2/R_1)}{A}} \tag{2.11}$$

Observe that the denominator is identical to that for the case of the inverting configuration (Eq. 2.5). This is no coincidence; it is a result of the fact that both the inverting and the noninverting configurations have the same feedback loop, which can be readily seen if the input signal source is eliminated (i.e., short-circuited). The numerators, however, are different, for the numerator gives the ideal or nominal closed-loop gain ($-R_2/R_1$ for the inverting configuration, and $1 + R_2/R_1$ for the noninverting configuration). Finally, we note (with reassurance) that the gain expression in Eq. (2.11) reduces to the ideal value for $A = \infty$. In fact, it approximates the ideal value for

$$A \gg 1 + \frac{R_2}{R_1}$$

This is the same condition as in the inverting configuration, except that here the quantity on the right-hand side is the nominal closed-loop gain. The expressions for the actual and ideal values of the closed-loop gain G in Eqs. (2.11) and (2.9), respectively, can be used to determine the percentage error in G resulting from the finite op-amp gain A as

$$\text{Percent gain error} = -\frac{1 + (R_2/R_1)}{A + 1 + (R_2/R_1)} \times 100 \tag{2.12}$$

Thus, as an example, if an op amp with an open-loop gain of 1000 is used to design a noninverting amplifier with a nominal closed-loop gain of 10, we would expect the closed-loop gain to be about 1% below the nominal value.

2.3.3 Input and Output Resistance

The gain of the noninverting configuration is positive—hence the name *noninverting*. The input impedance of this closed-loop amplifier is ideally infinite, since no current flows into the positive input terminal of the op amp. The output of the noninverting amplifier is taken at the terminals of the ideal voltage source $A(v_2 - v_1)$ (see the op-amp equivalent circuit in Fig. 2.3), and thus the output resistance of the noninverting configuration is zero.

2.3.4 The Voltage Follower

The property of high input impedance is a very desirable feature of the noninverting configuration. It enables using this circuit as a buffer amplifier to connect a source with a high

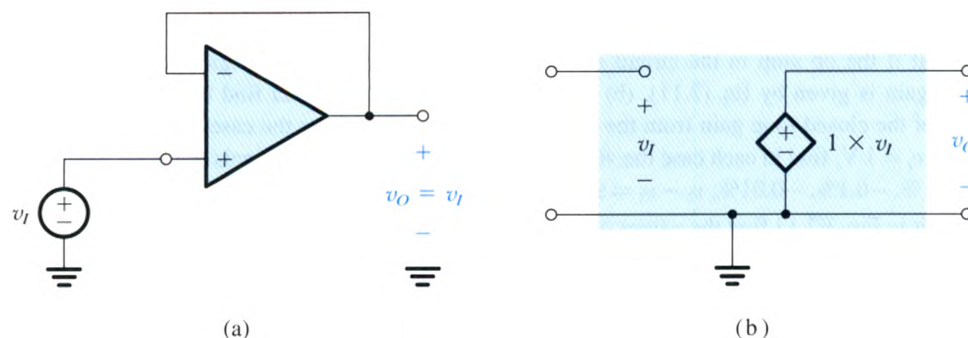


Figure 2.14 (a) The unity-gain buffer or follower amplifier. (b) Its equivalent circuit model.

impedance to a low-impedance load. We discussed the need for buffer amplifiers in Section 1.5. In many applications the buffer amplifier is not required to provide any voltage gain; rather, it is used mainly as an impedance transformer or a power amplifier. In such cases we may make $R_2 = 0$ and $R_1 = \infty$ to obtain the **unity-gain amplifier** shown in Fig. 2.14(a). This circuit is commonly referred to as a **voltage follower**, since the output voltage "follows" the input voltage. In the ideal case, $v_O = v_I, R_{in} = \infty$, $R_{out} = 0$, and the follower has the equivalent circuit shown in Fig. 2.14(b).

Since in the voltage-follower circuit the entire output is fed back to the inverting input, the circuit is said to have 100% negative feedback. The infinite gain of the op amp then acts to make $v_{Id} = 0$ and hence $v_O = v_I$. Observe that the circuit is elegant in its simplicity!

Since the noninverting configuration has a gain greater than or equal to unity, depending on the choice of R_2/R_1, some prefer to call it "a follower with gain."

EXERCISES

2.9 Use the superposition principle to find the output voltage of the circuit shown in Fig. E2.9.
Ans. $v_O = 6v_1 + 2v_2$

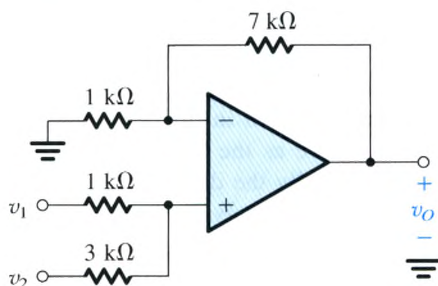


Figure E2.9

2.10 If in the circuit of Fig. E2.9 the 1-kΩ resistor is disconnected from ground and connected to a third signal source v_3, use superposition to determine v_O in terms of v_1, v_2, and v_3.
Ans. $v_O = 6v_1 + 2v_2 - 7v_3$

D2.11 Design a noninverting amplifier with a gain of 2. At the maximum output voltage of 10 V the current in the voltage divider is to be 10 μA.
Ans. $R_1 = R_2 = 0.5\ \text{M}\Omega$

2.12 (a) Show that if the op amp in the circuit of Fig. 2.12 has a finite open-loop gain A, then the closed-loop gain is given by Eq. (2.11). (b) For $R_1 = 1$ kΩ and $R_2 = 9$ kΩ find the percentage deviation ϵ of the closed-loop gain from the ideal value of $(1 + R_2/R_1)$ for the cases $A = 10^3, 10^4$, and 10^5. For $v_I = 1$ V, find in each case the voltage between the two input terminals of the op amp.
Ans. $\epsilon = -1\%, -0.1\%, -0.01\%$; $v_2 - v_1 = 9.9$ mV, 1 mV, 0.1 mV

2.13 For the circuit in Fig. E2.13 find the values of i_I, v_1, i_1, i_2, v_O, i_L, and i_O. Also find the voltage gain v_O/v_I, the current gain i_L/i_I, and the power gain P_L/P_I.
Ans. 0; 1 V; 1 mA; 1 mA; 10 V; 10 mA; 11 mA; 10 V/V (20 dB); ∞; ∞

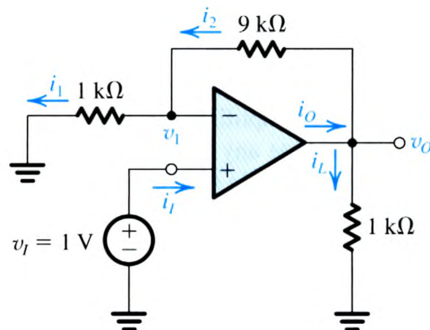


Figure E2.13

2.14 It is required to connect a transducer having an open-circuit voltage of 1 V and a source resistance of 1 MΩ to a load of 1-kΩ resistance. Find the load voltage if the connection is done (a) directly, and (b) through a unity-gain voltage follower.
Ans. (a) 1 mV; (b) 1 V

2.4 Difference Amplifiers

Having studied the two basic configurations of op-amp circuits together with some of their direct applications, we are now ready to consider a somewhat more involved but very important application. Specifically, we shall study the use of op amps to design difference or differential amplifiers.[2] A difference amplifier is one that responds only to the difference between the two signals applied at its input. The representation of signals in terms of their differential and common-mode components was given in Fig. 2.4. It is repeated here in Fig. 2.15 with slightly different symbols to serve as the input signals for the difference amplifiers we are about to design. Although ideally the difference amplifier will amplify only the differential input signal v_{Id} and reject completely the common-mode input signal v_{Icm}, practical circuits will have an output voltage v_O given by

$$v_O = A_d v_{Id} + A_{cm} v_{Icm} \tag{2.13}$$

where A_d denotes the amplifier differential gain and A_{cm} denotes its common-mode gain (ideally zero). The efficacy of a differential amplifier is measured by the degree of its

[2]The terms *difference* and *differential* are usually used to describe somewhat different amplifier types. For our purposes at this point, the distinction is not sufficiently significant. We will be more precise near the end of this section.

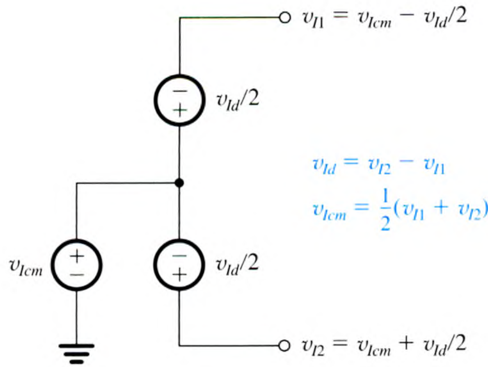


Figure 2.15 Representing the input signals to a differential amplifier in terms of their differential and common-mode components.

rejection of common-mode signals in preference to differential signals. This is usually quantified by the **common-mode rejection ratio (CMRR)**,

$$\text{CMRR} = 20\log\frac{|A_d|}{|A_{cm}|} \tag{2.14}$$

The need for difference amplifiers arises frequently in the design of electronic systems, especially those employed in instrumentation. As a common example, consider a transducer providing a small (e.g., 1 mV) signal between its two output terminals while each of the two wires leading from the transducer terminals to the measuring instrument may have a large interference signal (e.g., 1 V) relative to the circuit ground. The instrument front end obviously needs a difference amplifier.

Before we proceed any further we should address a question that the reader might have: The op amp is itself a difference amplifier; why not just use an op amp? The answer is that the very high (ideally infinite) gain of the op amp makes it impossible to use by itself. Rather, as we did before, we have to devise an appropriate feedback network to connect to the op amp to create a circuit whose closed-loop gain is finite, predictable, and stable.

2.4.1 A Single-Op-Amp Difference Amplifier

Our first attempt at designing a difference amplifier is motivated by the observation that the gain of the noninverting amplifier configuration is positive, $(1 + R_2/R_1)$, while that of the inverting configuration is negative, $(-R_2/R_1)$. Combining the two configurations while ensuring that the two gain magnitudes are equal will therefore amplify the difference between two input signals. This requires attenuation of the positive input to reduce the positive path gain from $(1+R_2/R_1)$ to (R_2/R_1). A voltage divider (R_3, R_4) is used for this purpose, resulting in the circuit shown in Fig. 2.16. The proper ratio of the voltage divider can be determined as follows:

$$\frac{R_4}{R_4+R_3}\left(1+\frac{R_2}{R_1}\right) = \frac{R_2}{R_1} \Rightarrow \frac{R_4}{R_4+R_3} = \frac{R_2}{R_2+R_1}$$

This can be satisfied by selecting

$$\frac{R_4}{R_3} = \frac{R_2}{R_1} \tag{2.15}$$

Let's now step back and verify that the circuit in Fig. 2.16 with R_3 and R_4 selected according to Eq. (2.15) does in fact function as a difference amplifier. Specifically, we wish to

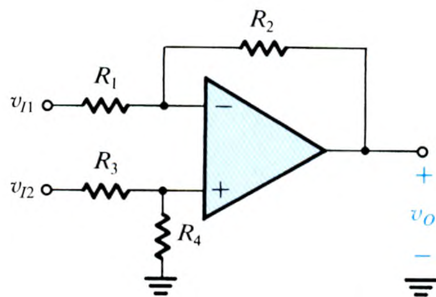


Figure 2.16 A difference amplifier.

determine the output voltage v_O in terms of v_{I1} and v_{I2}. To that end, note that the circuit is linear, and thus we can use superposition.

To apply superposition, we first reduce v_{I2} to zero—that is, ground the terminal to which v_{I2} is applied—and then find the corresponding output voltage, which will be due entirely to v_{I1}. We denote this output voltage v_{O1}. Its value may be found from the circuit in Fig. 2.17(a), which we recognize as that of the inverting configuration. The existence of R_3 and R_4 does not affect the gain expression, since no current flows through either of them. Thus,

$$v_{O1} = -\frac{R_2}{R_1} v_{I1}$$

Next, we reduce v_{I1} to zero and evaluate the corresponding output voltage v_{O2}. The circuit will now take the form shown in Fig. 2.17(b), which we recognize as the noninverting configuration with an additional voltage divider, made up of R_3 and R_4, connected to the input v_{I2}. The output voltage v_{O2} is therefore given by

$$v_{O2} = v_{I2} \frac{R_4}{R_3 + R_4}\left(1 + \frac{R_2}{R_1}\right) = \frac{R_2}{R_1} v_{I2}$$

where we have utilized Eq. (2.15).

The superposition principle tells us that the output voltage v_O is equal to the sum of v_{O1} and v_{O2}. Thus we have

$$v_O = \frac{R_2}{R_1}(v_{I2} - v_{I1}) = \frac{R_2}{R_1} v_{Id} \tag{2.16}$$

Thus, as expected, the circuit acts as a difference amplifier with a differential gain A_d of

$$A_d = \frac{R_2}{R_1} \tag{2.17}$$

Of course this is predicated on the op amp being ideal and on the selection of R_3 and R_4 so that their ratio matches that of R_1 and R_2 (Eq. 2.15). To make this matching requirement easier to satisfy, we usually select

$$R_3 = R_1 \quad \text{and} \quad R_4 = R_2$$

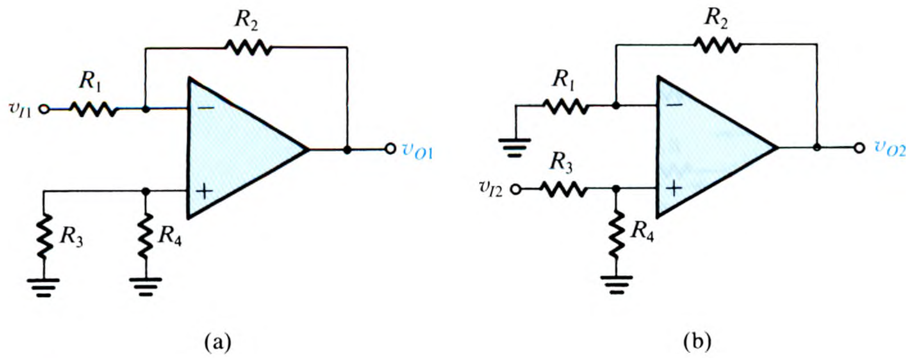


Figure 2.17 Application of superposition to the analysis of the circuit of Fig. 2.16.

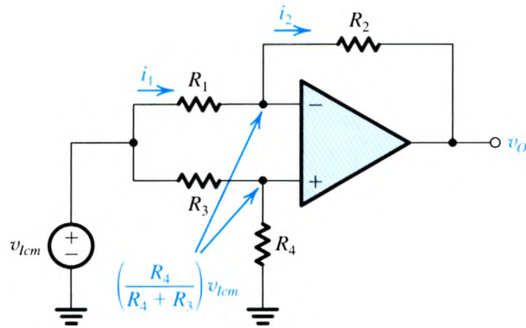


Figure 2.18 Analysis of the difference amplifier to determine its common-mode gain $A_{cm} \equiv v_O/v_{Icm}$.

Let's next consider the circuit with only a common-mode signal applied at the input, as shown in Fig. 2.18. The figure also shows some of the analysis steps. Thus,

$$
\begin{aligned}
i_1 &= \frac{1}{R_1}\left[v_{Icm} - \frac{R_4}{R_4+R_3}v_{Icm}\right] \\
&= v_{Icm}\frac{R_3}{R_4+R_3}\frac{1}{R_1}
\end{aligned}
\tag{2.18}
$$

The output voltage can now be found from

$$v_O = \frac{R_4}{R_4+R_3}v_{Icm} - i_2 R_2$$

Substituting $i_2 = i_1$ and for i_1 from Eq. (2.18),

$$
\begin{aligned}
v_O &= \frac{R_4}{R_4+R_3}v_{Icm} - \frac{R_2}{R_1}\frac{R_3}{R_4+R_3}v_{Icm} \\
&= \frac{R_4}{R_4+R_3}\left(1 - \frac{R_2}{R_1}\frac{R_3}{R_4}\right)v_{Icm}
\end{aligned}
$$

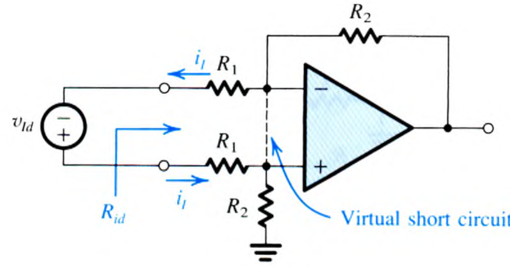


Figure 2.19 Finding the input resistance of the difference amplifier for the case $R_3 = R_1$ and $R_4 = R_2$.

Thus,

$$A_{cm} \equiv \frac{v_O}{v_{Icm}} = \left(\frac{R_4}{R_4 + R_3}\right)\left(1 - \frac{R_2}{R_1}\frac{R_3}{R_4}\right) \tag{2.19}$$

For the design with the resistor ratios selected according to Eq. (2.15), we obtain

$$A_{cm} = 0$$

as expected. Note, however, that any mismatch in the resistance ratios can make A_{cm} nonzero, and hence CMRR finite.

In addition to rejecting common-mode signals, a difference amplifier is usually required to have a high input resistance. To find the input resistance between the two input terminals (i.e., the resistance seen by v_{Id}), called the **differential input resistance** R_{id}, consider Fig. 2.19. Here we have assumed that the resistors are selected so that

$$R_3 = R_1 \quad \text{and} \quad R_4 = R_2$$

Now

$$R_{id} \equiv \frac{v_{Id}}{i_I}$$

Since the two input terminals of the op amp track each other in potential, we may write a loop equation and obtain

$$v_{Id} = R_1 i_I + 0 + R_1 i_I$$

Thus,

$$R_{id} = 2R_1 \tag{2.20}$$

Note that if the amplifier is required to have a large differential gain (R_2/R_1), then R_1 will need to be relatively small and the input resistance correspondingly low, a drawback of this circuit. Another drawback of the circuit is that it is not easy to vary the differential gain of the amplifier. Both drawbacks are overcome in the instrumentation amplifier discussed next.

EXERCISES

2.15 Consider the difference-amplifier circuit of Fig. 2.16 for the case $R_1 = R_3 = 2\text{ k}\Omega$ and $R_2 = R_4 = 200\text{ k}\Omega$. (a) Find the value of the differential gain A_d. (b) Find the value of the differential input resistance R_{id} and the output resistance R_o. (c) If the resistors have 1% tolerance (i.e., each can be within ±1% of its nominal value), use Eq. (2.19) to find the worst-case common-mode gain A_{cm} and hence the corresponding value of CMRR.

Ans. (a) 100 V/V (40 dB); (b) 4 kΩ, 0 Ω; (c) 0.04 V/V, 68 dB

D2.16 Find values for the resistances in the circuit of Fig. 2.16 so that the circuit behaves as a difference amplifier with an input resistance of 20 kΩ and a gain of 10.

Ans. $R_1 = R_3 = 10\text{ k}\Omega$; $R_2 = R_4 = 100\text{ k}\Omega$

2.4.2 A Superior Circuit: The Instrumentation Amplifier

The low-input-resistance problem of the difference amplifier of Fig. 2.16 can be solved by using voltage followers to buffer the two input terminals; that is, a voltage follower of the type in Fig. 2.14 is connected between each input terminal and the corresponding input terminal of the difference amplifier. However, if we are going to use two additional op amps, we should ask the question: Can we get some voltage gain from them? It is especially interesting that we can achieve this without compromising the high input resistance simply by using noninverting stages rather than unity-gain followers. Achieving some or indeed the bulk of

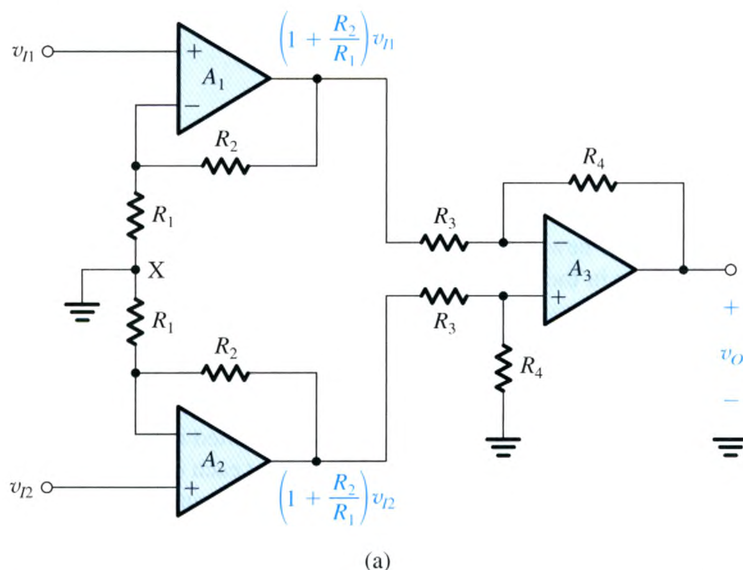


Figure 2.20 A popular circuit for an instrumentation amplifier. **(a)** Initial approach to the circuit. **(b)** The circuit in **(a)** with the connection between node X and ground removed and the two resistors R_1 and R_1 lumped together. This simple wiring change dramatically improves performance. **(c)** Analysis of the circuit in **(b)** assuming ideal op amps.

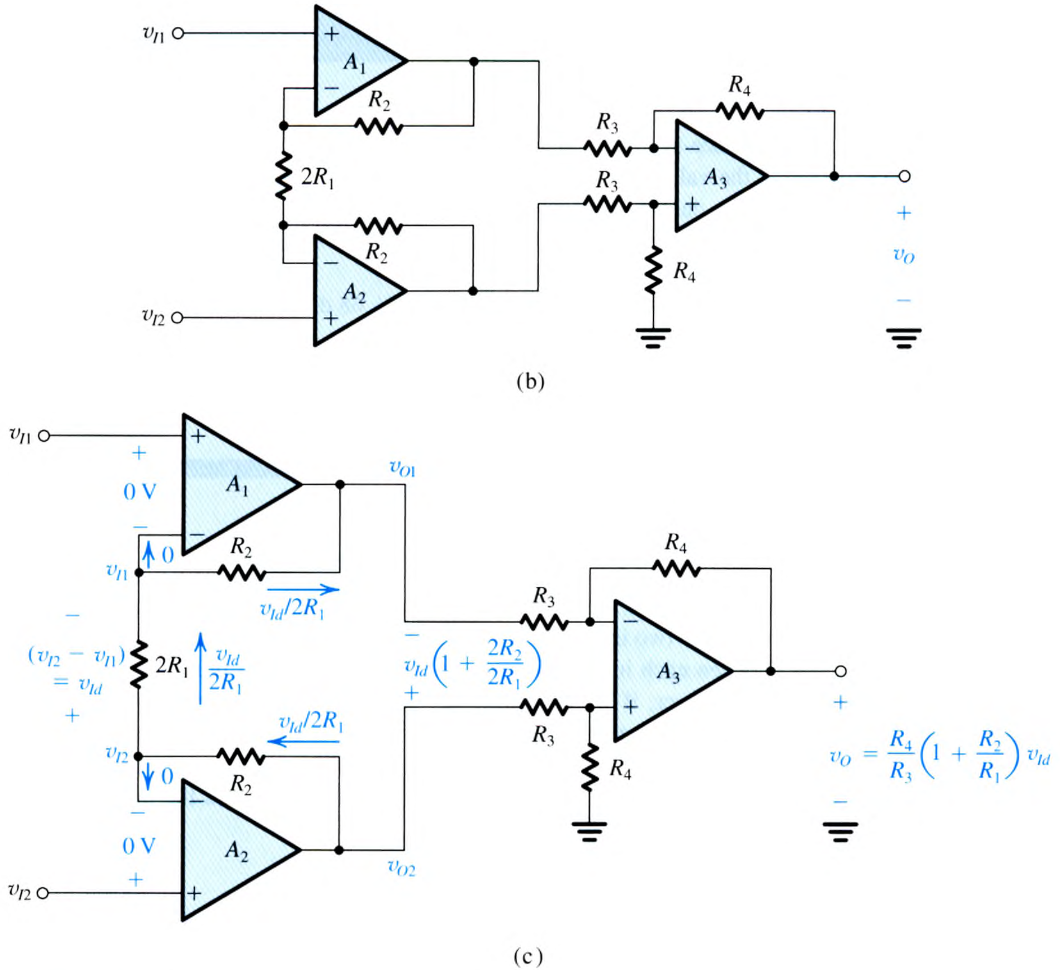


Figure 2.20 *continued*

the required gain in this new first stage of the differential amplifier eases the burden on the difference amplifier in the second stage, leaving it to its main task of implementing the differencing function and thus rejecting common-mode signals.

The resulting circuit is shown in Fig. 2.20(a). It consists of two stages in cascade. The first stage is formed by op amps A_1 and A_2 and their associated resistors, and the second stage is the by-now-familiar difference amplifier formed by op amp A_3 and its four associated resistors. Observe that as we set out to do, each of A_1 and A_2 is connected in the noninverting configuration and thus realizes a gain of $(1+R_2/R_1)$. It follows that each of v_{I1} and v_{I2} is amplified by this factor, and the resulting amplified signals appear at the outputs of A_1 and A_2, respectively.

The difference amplifier in the second stage operates on the difference signal $(1+R_2/R_1)(v_{I2}-v_{I1}) = (1+R_2/R_1)v_{Id}$ and provides at its output

$$v_O = \frac{R_4}{R_3}\left(1+\frac{R_2}{R_1}\right)v_{Id}$$

Thus the differential gain realized is

$$A_d = \left(\frac{R_4}{R_3}\right)\left(1 + \frac{R_2}{R_1}\right) \tag{2.21}$$

The common-mode gain will be zero because of the differencing action of the second-stage amplifier.

The circuit in Fig. 2.20(a) has the advantage of very high (ideally infinite) input resistance and high differential gain. Also, provided A_1 and A_2 and their corresponding resistors are matched, the two signal paths are symmetric—a definite advantage in the design of a differential amplifier. The circuit, however, has three major disadvantages:

1. The input common-mode signal v_{Icm} is amplified in the first stage by a gain equal to that experienced by the differential signal v_{Id}. This is a very serious issue, for it could result in the signals at the outputs of A_1 and A_3 being of such large magnitudes that the op amps saturate (more on op-amp saturation in Section 2.8). But even if the op amps do not saturate, the difference amplifier of the second stage will now have to deal with much larger common-mode signals, with the result that the CMRR of the overall amplifier will inevitably be reduced.
2. The two amplifier channels in the first stage have to be perfectly matched, otherwise a common-mode input signal may appear differentially between their two outputs. Such a signal would get amplified by the difference amplifier in the second stage.
3. To vary the differential gain A_d, two resistors have to be varied simultaneously, say the two resistors labeled R_1. At each gain setting the two resistors have to be perfectly matched: a difficult task.

All three problems can be solved with a very simple wiring change: Simply disconnect the node between the two resistors labeled R_1, node X, from ground. The circuit with this small but functionally profound change is redrawn in Fig. 2.20(b), where we have lumped the two resistors (R_1 and R_1) together into a single resistor ($2R_1$).

Analysis of the circuit in Fig. 2.20(b), assuming ideal op amps, is straightforward, as is illustrated in Fig. 2.20(c). The key point is that the virtual short circuits at the inputs of op amps A_1 and A_2 cause the input voltages v_{I1} and v_{I2} to appear at the two terminals of resistor ($2R_1$). Thus the differential input voltage $v_{I2} - v_{I1} \equiv v_{Id}$ appears across $2R_1$ and causes a current $i = v_{Id}/2R_1$ to flow through $2R_1$ and the two resistors labeled R_2. This current in turn produces a voltage difference between the output terminals of A_1 and A_2 given by

$$v_{O2} - v_{O1} = \left(1 + \frac{2R_2}{2R_1}\right)v_{Id}$$

The difference amplifier formed by op amp A_3 and its associated resistors senses the voltage difference ($v_{O2} - v_{O1}$) and provides a proportional output voltage v_O:

$$v_O = \frac{R_4}{R_3}(v_{O2} - v_{O1})$$

$$= \frac{R_4}{R_3}\left(1 + \frac{R_2}{R_1}\right)v_{Id}$$

Thus the overall differential voltage-gain is given by

$$A_d \equiv \frac{v_O}{v_{Id}} = \frac{R_4}{R_3}\left(1 + \frac{R_2}{R_1}\right) \tag{2.22}$$

Observe that proper differential operation does *not* depend on the matching of the two resistors labeled R_2. Indeed, if one of the two is of different value, say R_2', the expression for A_d becomes

$$A_d = \frac{R_4}{R_3}\left(1 + \frac{R_2 + R_2'}{2R_1}\right) \tag{2.23}$$

Consider next what happens when the two input terminals are connected together to a common-mode input voltage v_{Icm}. It is easy to see that an equal voltage appears at the negative input terminals of A_1 and A_2, causing the current through $2R_1$ to be zero. Thus there will be no current flowing in the R_2 resistors, and the voltages at the output terminals of A_1 and A_2 will be equal to the input (i.e., v_{Icm}). Thus the first stage no longer amplifies v_{Icm}; it simply propagates v_{Icm} to its two output terminals, where they are subtracted to produce a zero common-mode output by A_3. The difference amplifier in the second stage, however, now has a much improved situation at its input: The difference signal has been amplified by $(1 + R_2/R_1)$ while the common-mode voltage remained unchanged.

Finally, we observe from the expression in Eq. (2.22) that the gain can be varied by changing only one resistor, $2R_1$. We conclude that this is an excellent differential amplifier circuit and is widely employed as an instrumentation amplifier, that is, as the input amplifier used in a variety of electronic instruments.

INTEGRATED INSTRUMENTATION AMPLIFIERS

The conventional combination of three op amps and a number of precision resistors to form an instrumentation amplifier is an extremely powerful tool for the design of instruments for many applications. While the earliest applications used separate op amps and discrete resistors, fully integrated versions incorporating most required components in a single integrated-circuit package are increasingly available from many manufacturers. Low-power versions of these units are extremely important in the design of portable, wearable, and implantable medical monitoring devices, such as wristband activity monitors.

Example 2.3

Design the instrumentation amplifier circuit in Fig. 2.20(b) to provide a gain that can be varied over the range of 2 to 1000 utilizing a 100-kΩ variable resistance (a potentiometer, or "pot" for short).

Solution

It is usually preferable to obtain all the required gain in the first stage, leaving the second stage to perform the task of taking the difference between the outputs of the first stage and thereby rejecting the common-mode signal. In other words, the second stage is usually designed for a gain of 1. Adopting this approach, we select all the second-stage resistors to be equal to a practically convenient value, say 10 kΩ. The problem then reduces to designing the first stage to realize a gain adjustable over the range of 2 to 1000. Implementing $2R_1$ as the series combination of a fixed resistor R_{1f} and the variable resistor R_{1v}

obtained using the 100-kΩ pot (Fig. 2.21), we can write

$$1+\frac{2R_2}{R_{1f}+R_{1v}}=2 \text{ to } 1000$$

Thus,

$$1+\frac{2R_2}{R_{1f}}=1000$$

and

$$1+\frac{2R_2}{R_{1f}+100\text{ k}\Omega}=2$$

These two equations yield $R_{1f}=100.2\ \Omega$ and $R_2=50.050\text{ k}\Omega$. Other practical values may be selected; for instance, selecting $R_{1f}=100\ \Omega$ and $R_2=49.9\text{ k}\Omega$ (both values are available as standard 1%-tolerance metal-film resistors; see Appendix J) results in a gain covering approximately the required range.

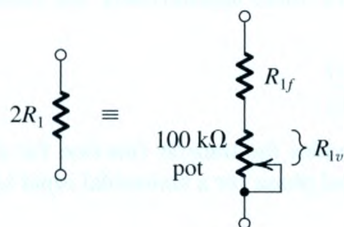


Figure 2.21 To make the gain of the circuit in Fig. 2.20(b) variable, $2R_1$ is implemented as the series combination of a fixed resistor R_{1f} and a variable resistor R_{1v}. Resistor R_{1f} ensures that the maximum available gain is limited.

EXERCISE

2.17 Consider the instrumentation amplifier of Fig. 2.20(b) with a common-mode input voltage of +5 V (dc) and a differential input signal of 10-mV-peak sine wave. Let $(2R_1)=1\text{ k}\Omega$, $R_2=0.5\text{ M}\Omega$, and $R_3=R_4=10\text{ k}\Omega$. Find the voltage at every node in the circuit.

Ans. $v_{I1}=5-0.005\sin\omega t$; $v_{I2}=5+0.005\sin\omega t$; v_-(op amp A_1)$=5-0.005\sin\omega t$; v_-(op amp A_2)$=5+0.005\sin\omega t$; $v_{O1}=5-5.005\sin\omega t$; $v_{O2}=5+5.005\sin\omega t$; $v_-(A_3)=v_+(A_3)=2.5+2.5025\sin\omega t$; $v_O=10.01\sin\omega t$ (all in volts)

2.5 Integrators and Differentiators

The op-amp circuit applications we have studied thus far utilized resistors in the op-amp feedback path and in connecting the signal source to the circuit, that is, in the feed-in path. As a result, circuit operation has been (ideally) independent of frequency. By allowing the

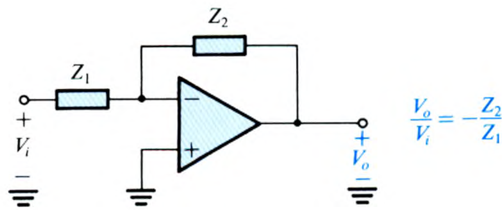


Figure 2.22 The inverting configuration with general impedances in the feedback and the feed-in paths.

use of capacitors together with resistors in the feedback and feed-in paths of op-amp circuits, we open the door to a very wide range of useful and exciting applications of the op amp. We begin our study of op-amp–RC circuits by considering two basic applications, namely, signal integrators and differentiators.[3]

2.5.1 The Inverting Configuration with General Impedances

To begin with, consider the inverting closed-loop configuration with impedances $Z_1(s)$ and $Z_2(s)$ replacing resistors R_1 and R_2, respectively. The resulting circuit is shown in Fig. 2.22 and, for an ideal op amp, has the closed-loop gain or, more appropriately, the closed-loop transfer function

$$\frac{V_o(s)}{V_i(s)} = -\frac{Z_2(s)}{Z_1(s)} \tag{2.24}$$

As explained in Section 1.6, replacing s by $j\omega$ provides the transfer function for physical frequencies ω, that is, the transmission magnitude and phase for a sinusoidal input signal of frequency ω.

EARLY OP AMPS AND ANALOG COMPUTATION

In 1941, Karl D. Swartzel Jr. of Bell Labs patented "the summing amplifier," a high-gain dc inverting amplifier, intended to be used with negative feedback. This precursor of the op amp used three vacuum tubes (the predecessor of the transistor) and ±350-V power supplies to achieve a gain of 90 dB. Though lacking a differential input, it provided the usual applications of summation, integration, and general filtering using convenient passive resistive and capacitive components.

Soon after (1942), Loebe Julie, working with Professor John R. Regazzini at Columbia University, created a differential version, still using vacuum tubes. During World War II, these units were used extensively to provide analog computational functions in association with radar-directed antiaircraft firing control involving aircraft speed projection.

In the early 1950s, driven by the demonstrated wartime success of op-amp-based computation, general-purpose commercial systems called "analog computers" began to appear. They consisted of a few dozen op amps and associated passive components, including potentiometers; the interconnections required for programming were achieved with plug boards. These computers were used to solve differential equations.

Today, analog integrated circuits are being used to accelerate the computations required for machine learning.

[3]At this point, a review of Section 1.6 would be helpful. Also, an important fact to remember: Passing a constant current I through a capacitor C for a time t causes a charge of It to accumulate on the capacitor. Thus the capacitor voltage changes by $\Delta V = \Delta Q/C = It/C$; that is, the capacitor voltage increases linearly with time.

Example 2.4

For the circuit in Fig. 2.23, derive an expression for the transfer function $V_o(s)/V_i(s)$. Show that the transfer function is that of a low-pass STC circuit. By expressing the transfer function in the standard form shown in Table 1.2, find the dc gain and the 3-dB frequency. Design the circuit to obtain a dc gain of 40 dB, a 3-dB frequency of 1 kHz, and an input resistance of 1 kΩ. At what frequency does the magnitude of transmission become unity? What is the phase angle at this frequency?

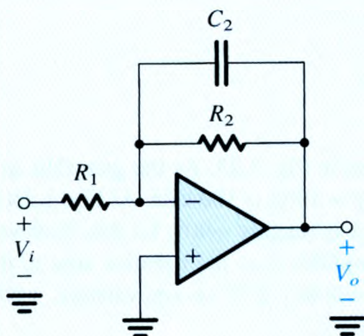


Figure 2.23 Circuit for Example 2.4.

Solution

To obtain the transfer function of the circuit in Fig. 2.23, we substitute in Eq. (2.24), $Z_1 = R_1$ and $Z_2 = R_2 \,\|\, (1/sC_2)$. Since Z_2 is the parallel connection of two components, it is more convenient to work in terms of Y_2; that is, we use the following alternative form of the transfer function:

$$\frac{V_o(s)}{V_i(s)} = -\frac{1}{Z_1(s)Y_2(s)}$$

and substitute $Z_1 = R_1$ and $Y_2(s) = (1/R_2) + sC_2$ to obtain

$$\frac{V_o(s)}{V_i(s)} = -\frac{1}{\dfrac{R_1}{R_2} + sC_2R_1}$$

This transfer function is of first order, has a finite dc gain (at $s = 0$, $V_o/V_i = -R_2/R_1$), and has zero gain at infinite frequency. Thus it is the transfer function of a low-pass STC network and can be expressed in the standard form of Table 1.2 as follows:

$$\frac{V_o(s)}{V_i(s)} = \frac{-R_2/R_1}{1 + sC_2R_2}$$

from which we find the dc gain K to be

$$K = -\frac{R_2}{R_1}$$

and the 3-dB frequency ω_0 as

$$\omega_0 = \frac{1}{C_2R_2}$$

Example 2.4 *continued*

We could have found all this from the circuit in Fig. 2.23 by inspection. Specifically, note that the capacitor behaves as an open circuit at dc; thus at dc the gain is simply $(-R_2/R_1)$. Furthermore, because there is a virtual ground at the inverting input terminal, the resistance seen by the capacitor is R_2, and thus the time constant of the STC network is C_2R_2.

Now to obtain a dc gain of 40 dB, that is, 100 V/V, we select $R_2/R_1 = 100$. For an input resistance of 1 kΩ, we select $R_1 = 1$ kΩ, and thus $R_2 = 100$ kΩ. Finally, for a 3-dB frequency $f_0 = 1$ kHz, we select C_2 from

$$2\pi \times 1 \times 10^3 = \frac{1}{C_2 \times 100 \times 10^3}$$

which yields $C_2 = 1.59$ nF.

The circuit has gain and phase Bode plots of the standard form in Fig. 1.23. As the gain falls at a rate of –20 dB/decade, it will reach 0 dB in two decades, that is, at $f = 100f_0 = 100$ kHz. As Fig. 1.23(b) shows, at such a frequency, which is much greater than f_0, the phase is roughly $-90°$. To this, however, we must add the 180° arising from the inverting nature of the amplifier (i.e., the negative sign in the transfer function expression). Thus at 100 kHz, the total phase shift will be $-270°$ or, equivalently, $+90°$.

2.5.2 The Inverting Integrator

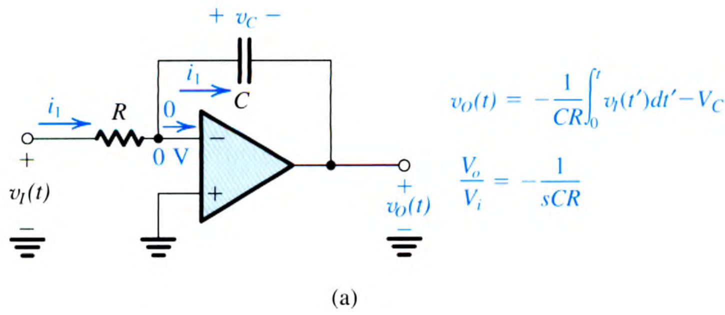
By placing a capacitor in the feedback path (i.e., in place of Z_2 in Fig. 2.22) and a resistor at the input (in place of Z_1), we obtain the circuit of Fig. 2.24(a). We shall now show that this circuit realizes the mathematical operation of integration. Let the input be a time-varying function $v_I(t)$. The virtual ground at the inverting op-amp input causes $v_I(t)$ to appear in effect across R, and thus the current $i_1(t)$ will be $v_I(t)/R$. This current flows through the capacitor C, causing charge to accumulate on C. If we assume that the circuit begins operation at time $t = 0$, then at an arbitrary time t the current $i_1(t)$ will have deposited on C a charge equal to $\int_0^t i_1(t')dt'$. Thus the capacitor voltage $v_C(t)$ will change by $\frac{1}{C}\int_0^t i_1(t')dt'$. If the initial voltage on C (at $t = 0$) is denoted V_C, then

$$v_C(t) = V_C + \frac{1}{C}\int_0^t i_1(t')dt'$$

Now the output voltage $v_O(t) = -v_C(t)$; thus,

$$v_O(t) = -\frac{1}{CR}\int_0^t v_I(t')dt' - V_C \quad (2.25)$$

Thus the circuit provides an output voltage that is proportional to the time integral of the input, with V_C being the initial condition of integration and CR the **integrator time constant**. Note that, as expected, there is a negative sign attached to the output voltage, and thus this integrator circuit is said to be an **inverting integrator**. It is also known as a **Miller integrator** after an early worker in this field.



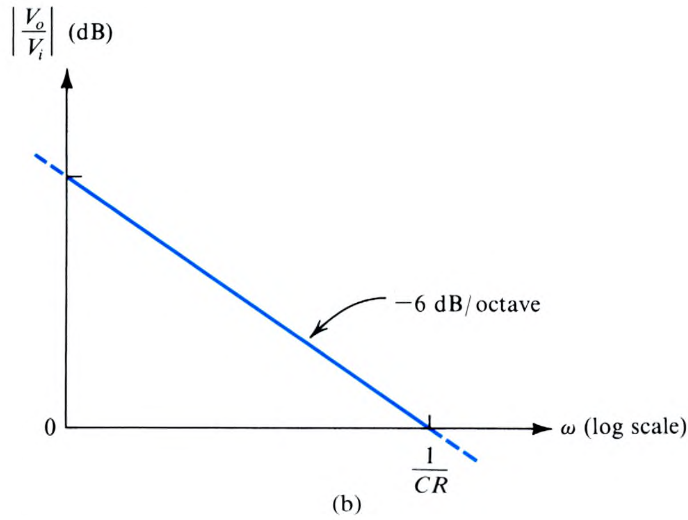


Figure 2.24 **(a)** The Miller or inverting integrator. **(b)** Frequency response of the integrator.

The operation of the integrator circuit can be described alternatively in the frequency domain by substituting $Z_1(s) = R$ and $Z_2(s) = 1/sC$ in Eq. (2.24) to obtain the transfer function

$$\frac{V_o(s)}{V_i(s)} = -\frac{1}{sCR} \tag{2.26}$$

For physical frequencies, $s = j\omega$ and

$$\frac{V_o(j\omega)}{V_i(j\omega)} = -\frac{1}{j\omega CR} \tag{2.27}$$

Thus the integrator transfer function has magnitude

$$\left|\frac{V_o}{V_i}\right| = \frac{1}{\omega CR} \tag{2.28}$$

and phase

$$\phi = +90^\circ \tag{2.29}$$

The Bode plot for the integrator magnitude response can be obtained by noting from Eq. (2.28) that as ω doubles (increases by an octave) the magnitude is halved (decreased

by 6 dB). Thus the Bode plot is a straight line of slope –6 dB/octave (or, equivalently, –20 dB/decade). This line (shown in Fig. 2.24b) intercepts the 0-dB line at the frequency that makes $|V_o/V_i| = 1$, which from Eq. (2.28) is

$$\omega_{\text{int}} = \frac{1}{CR} \tag{2.30}$$

The frequency ω_{int} is known as the **integrator frequency** and is simply the inverse of the integrator time constant.

Example 2.5

An integrator is commonly used to amplify signals from a photodetector (measuring light intensity). In Fig. 2.25, the photodetector is represented by the current source i_S producing a 10-kHz sinusoid having 10-μ*A* peak amplitude and a source resistance of 5 kΩ. Find the value of *C* and the integrator frequency ω_{int} so that the output at v_O will have 200-mV peak amplitude.

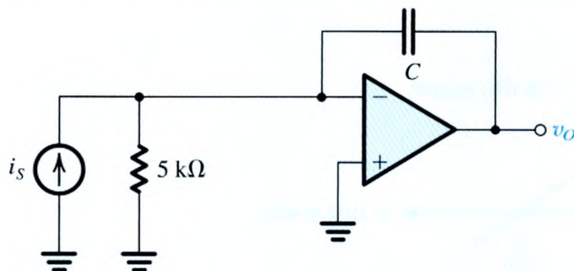


Figure 2.25 The integrator with input current source (Example 2.5).

Solution

Replacing the current source by its Thévenin equivalent results precisely in Fig. 2.24(a) with an ideal 10-kHz sinusoidal voltage source connected at the input, v_I, having 10 μA × 5 kΩ = 50 mV peak amplitude, and $R = 5$ kΩ. Using Eq. (2.28),

$$\left|\frac{200\text{ mV}}{50\text{ mV}}\right| = \frac{1}{2\pi \times 10\text{ kHz} \times C \times 5\text{ k}\Omega}$$

$$\Rightarrow C = \frac{1}{4 \times 2\pi \times 10\text{ kHz} \times 5\text{ k}\Omega} = 0.8\text{ nF}$$

Substitution into Eq. (2.30) results in

$$\omega_{int} = \frac{1}{0.8\text{ nF} \times 5\text{ k}\Omega} = 2\pi \times 40\text{ krad/s} \text{ or } f_{int} = 40\text{ kHz}$$

Comparing the frequency response of the integrator to that of an STC low-pass network indicates that the integrator behaves as a low-pass filter with a corner frequency of zero. Observe also that at $\omega = 0$, the magnitude of the integrator transfer function is infinite. This

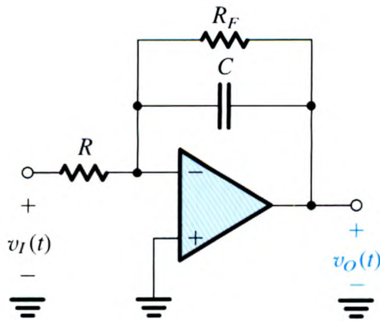


Figure 2.26 The Miller integrator with a large resistance R_F connected in parallel with C in order to provide negative feedback and hence finite gain at dc.

indicates that at dc the op amp is operating with an open loop. This should also be obvious from the integrator circuit itself. Reference to Fig. 2.24(a) shows that the feedback element is a capacitor, and thus at dc, where the capacitor behaves as an open circuit, there is no negative feedback! This is a very significant observation and one that indicates a source of problems with the integrator circuit: Any tiny dc component in the input signal will theoretically produce an infinite output. Of course, no infinite output voltage results in practice; rather, the output of the amplifier saturates at a voltage close to the op-amp positive or negative power supply (L_+ or L_-), depending on the polarity of the input dc signal.

The dc problem of the integrator circuit can be alleviated by connecting a resistor R_F across the integrator capacitor C, as shown in Fig. 2.26, and thus the gain at dc will be $-R_F/R$ rather than infinite. Such a resistor provides a dc feedback path. Unfortunately, however, the integration is no longer ideal, and the lower the value of R_F, the less ideal the integrator circuit becomes. This is because R_F causes the frequency of the integrator pole to move from its ideal location at $\omega = 0$ to one determined by the corner frequency of the STC network (R_F, C). Specifically, the integrator transfer function becomes

$$\frac{V_o(s)}{V_i(s)} = -\frac{R_F/R}{1 + sCR_F}$$

as opposed to the ideal function of $-1/sCR$. The lower the value we select for R_F, the higher the corner frequency $(1/CR_F)$ will be and the more nonideal the integrator becomes. Thus selecting a value for R_F presents the designer with a trade-off between dc performance and signal performance. The effect of R_F on integrator performance is investigated further in Example 2.6.

Example 2.6

Find the output produced by a Miller integrator in response to an input pulse of 1-V height and 1-ms width [Fig. 2.27(a)]. Let $R = 10\ \text{k}\Omega$ and $C = 10\ \text{nF}$. If the integrator capacitor is shunted by a 1-MΩ resistor, how will the response be modified? The op amp is specified to saturate at ±13 V.

Example 2.6 *continued*

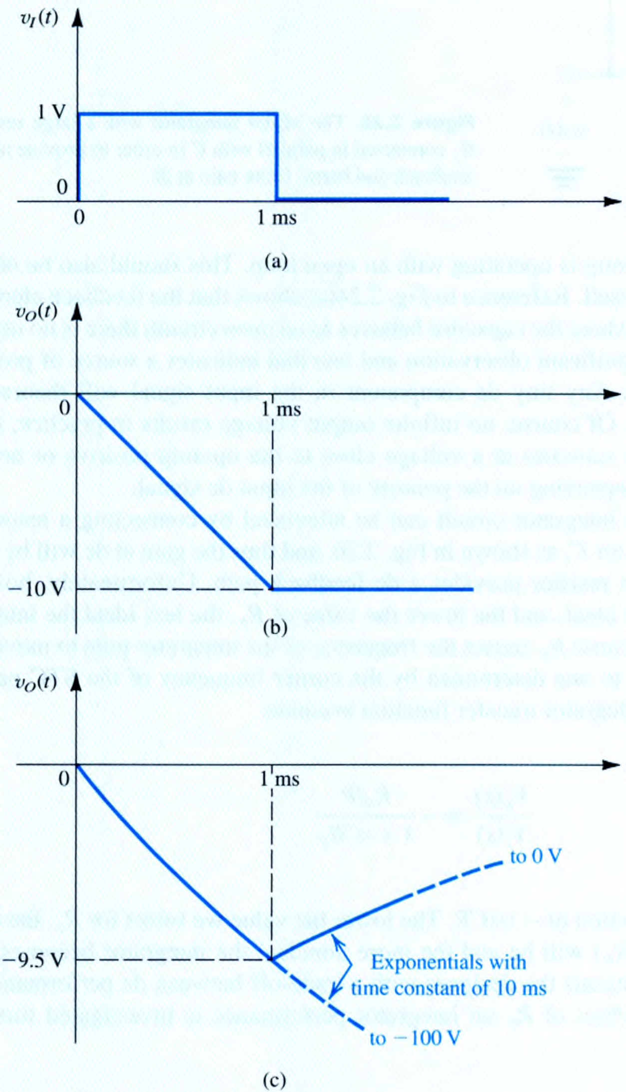


Figure 2.27 Waveforms for Example 2.6: **(a)** Input pulse. **(b)** Output linear ramp of ideal integrator with time constant of 0.1 ms. **(c)** Output exponential ramp with resistor R_F connected across integrator capacitor.

Solution

In response to a 1-V, 1-ms input pulse, the integrator output will be

$$v_O(t) = -\frac{1}{CR}\int_0^t 1\,dt', \qquad 0 \le t \le 1 \text{ ms}$$

where we have assumed that the initial voltage on the integrator capacitor is 0. For $C = 10\text{ nF}$ and $R = 10\text{ k}\Omega$, $CR = 0.1\text{ ms}$, and

$$v_O(t) = -10t, \qquad 0 \le t \le 1\text{ ms}$$

which is the linear ramp shown in Fig. 2.27(b). It reaches a magnitude of -10 V at $t = 1$ ms and remains constant thereafter.

That the output is a linear ramp should also be obvious from the fact that the 1-V input pulse produces a constant current through the capacitor of 1 V/10 kΩ = 0.1 mA. This constant current $I = 0.1$ mA supplies the capacitor with a charge It, and thus the capacitor voltage changes linearly as (It/C), resulting in $v_O = -(I/C)t$. It is worth remembering that charging a capacitor with a constant current produces a linear voltage across it.

Next consider the situation with resistor $R_F = 1\text{ M}\Omega$ connected across C. As before, the 1-V pulse will provide a constant current $I = 0.1$ mA. Now, however, this current is supplied to an STC network composed of R_F in parallel with C. Thus, the output will be an exponential heading toward -100 V with a time constant of $CR_F = 10 \times 10^{-9} \times 1 \times 10^6 = 10\text{ ms}$,

$$v_O(t) = -100(1 - e^{-t/10}), \qquad 0 \le t \le 1\text{ ms}$$

Of course, the exponential will be interrupted at the end of the pulse, that is, at $t = 1$ ms, and the output will reach the value

$$v_O(1\text{ ms}) = -100(1 - e^{-1/10}) = -9.5\text{ V}$$

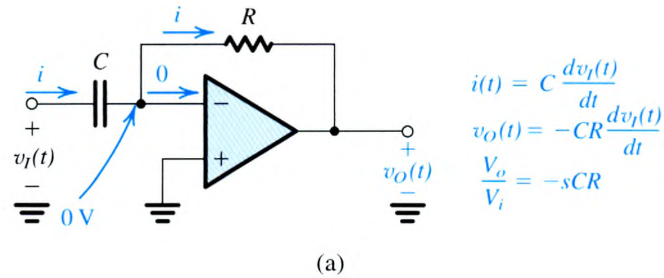
The output waveform is shown in Fig. 2.27(c), from which we see that including R_F causes the ramp to be slightly rounded such that the output reaches only -9.5 V, 0.5 V short of the ideal value of -10 V. Furthermore, for $t > 1$ ms, the capacitor discharges through R_F with the relatively long time constant of 10 ms. Finally, we note that op-amp saturation, specified to occur at ± 13 V, has no effect on the operation of this circuit.

The preceding example hints at an important application of integrators, namely, their use in providing triangular waveforms in response to square-wave inputs. This application is explored in Exercise 2.18. Integrators have many other applications, including their use in the design of filters (Chapter 14).

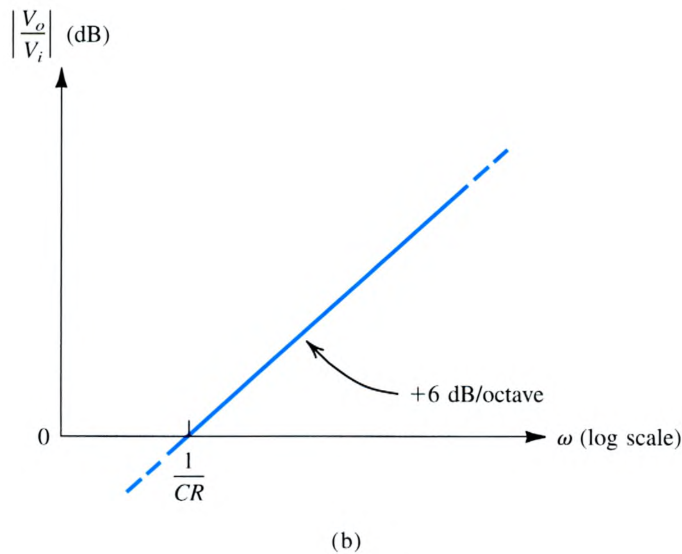
2.5.3 The Op-Amp Differentiator

Interchanging the location of the capacitor and resistor in the integrator circuit results in the circuit in Fig. 2.28(a), which performs differentiation of the input signal $v_I(t)$. The virtual ground at the op-amp inverting input terminal causes $v_I(t)$ to appear across the capacitor C. Thus the current through C will be $C(dv_I(t)/dt)$, which flows through the feedback resistor R resulting in the op-amp output voltage,

$$v_O(t) = -CR\frac{dv_I(t)}{dt} \tag{2.31}$$



(a)



(b)

Figure 2.28 **(a)** A differentiator. **(b)** Frequency response of a differentiator with a time constant CR.

The transfer function of the circuit can be found from Eq (2.24),

$$\frac{V_o(s)}{V_i(s)} = -sCR \tag{2.32}$$

from which the magnitude response is found by substituting $s = j\omega$,

$$\frac{V_o(j\omega)}{V_i(j\omega)} = -j\omega CR \Rightarrow \left|\frac{V_o}{V_i}\right| = \omega CR \tag{2.33}$$

Plotted in Fig. 2.28(b), the magnitude response increases at +6 dB/octave (+20 dB/decade) and exhibits large (ideally infinite) gain at high frequency. Unfortunately, the differentiator may experience a lot of high-frequency noise at the output and may suffer from instability (Chapter 11). Hence, in practice, it is usually necessary to connect a small-valued resistor in series with the capacitor.

EXERCISES

2.18 Consider a symmetrical square wave of 5-V peak-to-peak, 0 average, and 2-μs period applied to a Miller integrator. Find the value of the time constant CR such that the triangular waveform at the output has a 5-V peak-to-peak amplitude.
Ans. 0.5 μs

D2.19 Use an ideal op amp to design an inverting integrator with an input resistance of 10 kΩ and an integration time constant of 10^{-3} s. What is the gain magnitude and phase angle of this circuit at 10 rad/s and at 1 rad/s? What is the frequency at which the gain magnitude is unity?
Ans. $R = 10$ kΩ, $C = 0.1$ μF; at $\omega = 10$ rad/s: $|V_o/V_i| = 100$ V/V and $\phi = +90°$; at $\omega = 1$ rad/s: $|V_o/V_i| = 1000$ V/V and $\phi = +90°$; 1000 rad/s

D2.20 Design a differentiator to have a time constant of 10^{-2} s and an input capacitance of 0.01 μF. What is the gain magnitude and phase of this circuit at 10 rad/s, and at 10^3 rad/s? In order to limit the high-frequency gain of the differentiator circuit to 100, a resistor is added in series with the capacitor. Find the required resistor value.
Ans. $C = 0.01$ μF; $R = 1$ MΩ; at $\omega = 10$ rad/s: $|V_o/V_i| = 0.1$ V/V and $\phi = -90°$; at $\omega = 1000$ rad/s: $|V_o/V_i| = 10$ V/V and $\phi = -90°$; 10 kΩ

2.6 DC Imperfections

Thus far we have primarily considered the op amp to be ideal. Being thoroughly familiar with the characteristics of practical op amps and the effects of such characteristics on the performance of op-amp circuits will allow the designer to use the op amp intelligently, especially if the application at hand is not a straightforward one. The nonideal properties of op amps will, of course, limit the range of operation of the circuits analyzed in the previous examples.

In this and the next two sections we consider some of the important nonideal properties of the op amp.[4] We will treat one nonideality at a time, beginning in this section with the dc problems to which op amps are susceptible.

2.6.1 Offset Voltage

Because op amps are direct-coupled devices with large gains at dc, they are prone to dc problems. The first such problem is the dc offset voltage. To understand this problem consider the following *conceptual* experiment: If the two input terminals of the op amp are tied together and connected to ground, we will find that despite the fact that $v_{Id} = 0$, a finite dc voltage exists at the output. In fact, if the op amp has a high dc gain, the output will be at either the positive or negative saturation level. The op-amp output can be brought back to its ideal value of 0 V by connecting a dc voltage source of appropriate polarity and magnitude

[4] Real op amps have nonideal effects in addition to those discussed in this chapter. These include finite (nonzero) common-mode gain or, equivalently, noninfinite CMRR, noninfinite input resistance, and nonzero output resistance. The effect of these on the performance of most of the closed-loop circuits studied here is not very significant, and so their study will be postponed to later chapters (in particular, Chapters 9 and 13).

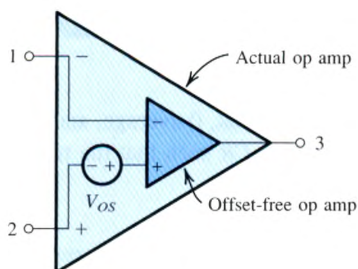


Figure 2.29 Circuit model for an op amp with input offset voltage V_{OS}.

between the two input terminals of the op amp. This external source balances out the input offset voltage of the op amp. It follows that the **input offset voltage** (V_{OS}) must be of equal magnitude and of opposite polarity to the voltage we applied externally.

The input offset voltage arises as a result of the unavoidable mismatches present in the input differential stage inside the op amp. In later chapters (in particular Chapters 9 and 13) we shall study this topic in detail. Here, however, our concern is to investigate the effect of V_{OS} on the operation of closed-loop op-amp circuits. To that end, we note that general-purpose op amps exhibit V_{OS} in the range of 1 mV to 5 mV. Also, the value of V_{OS} depends on temperature. The op-amp data sheets usually specify typical and maximum values for V_{OS} at room temperature as well as the temperature coefficient of V_{OS} (usually in μV/°C). They do not, however, specify the polarity of V_{OS} because the component mismatches that give rise to V_{OS} are obviously not known a priori; different units of the same op-amp type may exhibit either a positive or a negative V_{OS}.

To analyze the effect of V_{OS} on the operation of op-amp circuits, we need a circuit model for the op amp with input offset voltage. Such a model is shown in Fig. 2.29. It consists of a dc source of value V_{OS} placed in series with the positive input lead of an offset-free op amp. The justification for this model follows from the description above.

EXERCISE

2.21 Use the model of Fig. 2.29 to sketch the transfer characteristic v_O versus v_{Id} ($v_O \equiv v_3$ and $v_{Id} \equiv v_2 - v_1$) of an op amp having an open-loop dc gain $A_0 = 10^4$ V/V, output saturation levels of ±10 V, and V_{OS} of +5 mV.

Ans. See Fig. E2.21. Observe that true to its name, the input offset voltage causes an offset in the voltage-transfer characteristic; rather than passing through the origin it is now shifted to the left by V_{OS}.

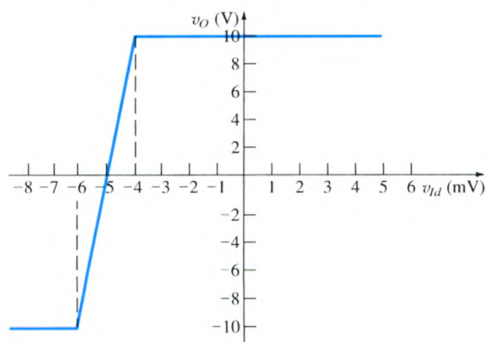


Figure E2.21 Transfer characteristic of an op amp with V_{OS} = 5 mV and output saturation levels ±10 V.

Analysis of op-amp circuits to determine the effect of the op-amp V_{OS} on their performance is straightforward: The input voltage signal source is short-circuited and the op amp is replaced with the model of Fig. 2.29. (Eliminating the input signal, done to simplify matters, is based on the principle of superposition.) Following this procedure, we find that both the inverting and the noninverting amplifier configurations result in the same circuit, that shown in Fig. 2.30, from which the output dc voltage due to V_{OS} is found to be

$$V_O = V_{OS}\left[1 + \frac{R_2}{R_1}\right] \tag{2.34}$$

This output dc voltage can have a large magnitude. For instance, a noninverting amplifier with a closed-loop gain of 1000, when constructed from an op amp with a 5-mV input offset voltage, will have a dc output voltage of +5 V or −5 V (depending on the polarity of V_{OS}) rather than the ideal value of 0 V. Now, when an input signal is applied to the amplifier, the corresponding signal output will be superimposed on the 5-V dc. Obviously then, the allowable signal swing at the output will be reduced. Even worse, if the signal to be amplified is dc, we would not know whether the output is due to V_{OS} or to the signal!

Some op amps are provided with two additional terminals to which a specified circuit can be connected to trim to zero the output dc voltage due to V_{OS}. Figure 2.31 shows such an arrangement that is typically used with general-purpose op amps. A potentiometer is connected between the offset-nulling terminals with the wiper of the potentiometer connected to the op-amp negative supply. Moving the potentiometer wiper introduces an imbalance that counteracts the asymmetry present in the internal op-amp circuitry and that gives rise to V_{OS}. We shall return to this point in the context of our study of the internal circuitry of op amps

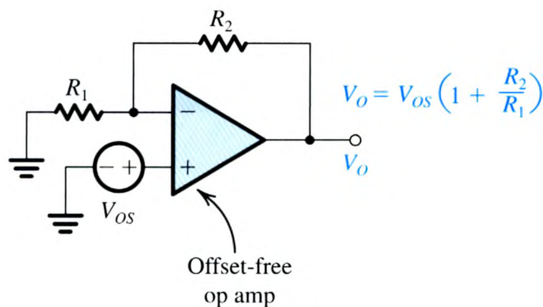


Figure 2.30 Evaluating the output dc offset voltage due to V_{OS} in a closed-loop amplifier.

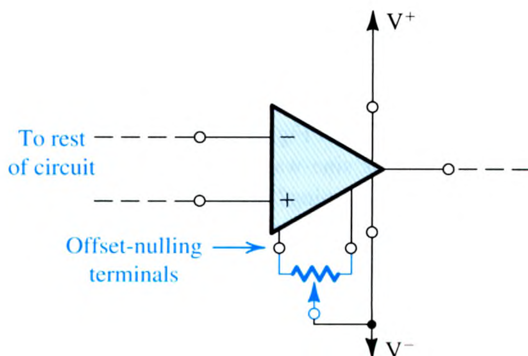


Figure 2.31 The output dc offset voltage of an op amp can be trimmed to zero by connecting a potentiometer to the two offset-nulling terminals. The wiper of the potentiometer is connected to the negative supply of the op amp.

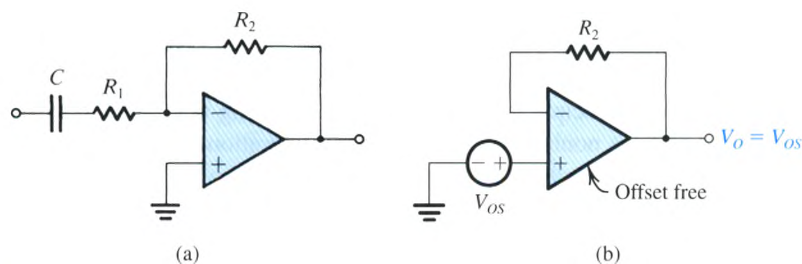


Figure 2.32 **(a)** A capacitively coupled inverting amplifier. **(b)** The equivalent circuit for determining its dc output offset voltage V_O.

in Chapter 13. It should be noted, however, that even though the dc output offset can be trimmed to zero, the problem remains of the variation (or drift) of V_{OS} with temperature.

One way to overcome the dc offset problem is by capacitively coupling the amplifier. This, however, will be possible only in applications where the closed-loop amplifier is not required to amplify dc or very-low-frequency signals. Figure 2.32(a) shows a capacitively coupled amplifier. Because of its infinite impedance at dc, the coupling capacitor will cause the gain to be zero at dc. As a result, the equivalent circuit for determining the dc output voltage resulting from the op-amp input offset voltage V_{OS} will be that shown in Fig. 2.32(b). Thus V_{OS} sees in effect a unity-gain voltage follower, and the dc output voltage V_O will be equal to V_{OS} rather than $V_{OS}(1 + R_2/R_1)$, which is the case without the coupling capacitor. As far as input signals are concerned, the coupling capacitor C forms together with R_1 an STC high-pass circuit with a corner frequency of $\omega_0 = 1/CR_1$. Thus the gain of the capacitively coupled amplifier will fall off at the low-frequency end [from a magnitude of $(1 + R_2/R_1)$ at high frequencies] and will be 3 dB down at ω_0.

EXERCISES

2.22 Consider an inverting amplifier with a nominal gain of 1000 constructed from an op amp with an input offset voltage of 3 mV and with output saturation levels of ±10 V. (a) What is (approximately) the peak sine-wave input signal that can be applied without output clipping? (b) If the effect of V_{OS} is nulled at room temperature (25°C), how large an input can one now apply if: (i) the circuit is to operate at a constant temperature? (ii) the circuit is to operate at a temperature in the range 0°C to 75°C and the temperature coefficient of V_{OS} is 10 μV/°C?
Ans. (a) 7 mV; (b) 10 mV, 9.5 mV

2.23 Consider the same amplifier as in Exercise 2.22—that is, an inverting amplifier with a nominal gain of 1000 constructed from an op amp with an input offset voltage of 3 mV and with output saturation levels of ±10 V—except here let the amplifier be capacitively coupled as in Fig. 2.32(a). (a) What is the dc offset voltage at the output, and what (approximately) is the peak sine-wave signal that can be applied at the input without output clipping? Is there a need for offset trimming? (b) If $R_1 = 1$ kΩ and $R_2 = 1$ MΩ, find the value of the coupling capacitor C_1 that will ensure that the gain will be greater than 57 dB down to 1 kHz.
Ans. (a) 3 mV, 10 mV, no need for offset trimming; (b) 0.16 μF

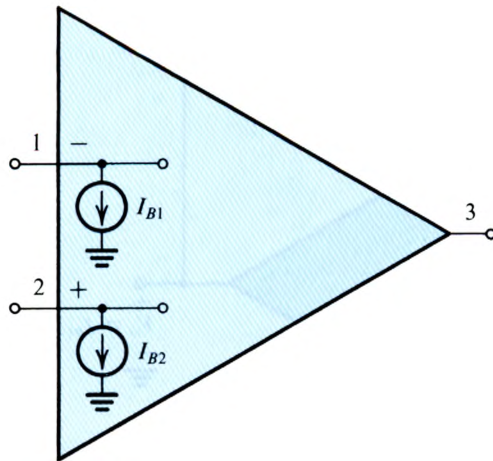


Figure 2.33 The op-amp input bias currents represented by two current sources I_{B1} and I_{B2}.

2.6.2 Input Bias and Offset Currents

The second dc problem encountered in op amps is illustrated in Fig. 2.33. In order for the op amp to operate, its two input terminals have to be supplied with dc currents, termed the **input bias currents.**[5] In Fig. 2.33 these two currents are represented by two current sources, I_{B1} and I_{B2}, connected to the two input terminals. It should be emphasized that the input bias currents are independent of the fact that a real op amp has finite (though large) input resistance (not shown in Fig. 2.33). The op-amp manufacturer usually specifies the average value of I_{B1} and I_{B2} as well as their expected difference. The average value I_B is called the **input bias current**,

$$I_B = \frac{I_{B1} + I_{B2}}{2}$$

and the difference is called the **input offset current** and is given by

$$I_{OS} = |I_{B1} - I_{B2}|$$

Typical values for general-purpose op amps that use bipolar transistors are $I_B = 100$ nA and $I_{OS} = 10$ nA.

We now wish to find the dc output voltage of the closed-loop amplifier due to the input bias currents. To do this we ground the signal source and obtain the circuit shown in Fig. 2.34 for both the inverting and noninverting configurations. As shown in Fig. 2.34, the output dc voltage is given by

$$V_O = I_{B1}R_2 \simeq I_B R_2 \tag{2.35}$$

This obviously places an upper limit on the value of R_2. Fortunately, however, there is a technique for reducing the value of the output dc voltage due to the input bias currents. The

[5] This is the case for op amps constructed using bipolar junction transistors (BJTs). Those using MOSFETs in the first (input) stage do not draw an appreciable input bias current; nevertheless, the input terminals should have continuous dc paths to ground. More on this in later chapters.

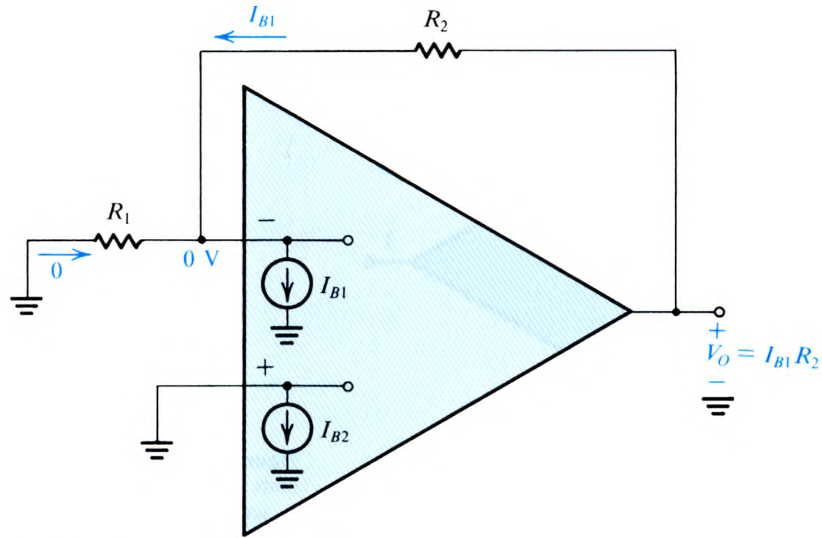


Figure 2.34 Analysis of the closed-loop amplifier, taking into account the input bias currents.

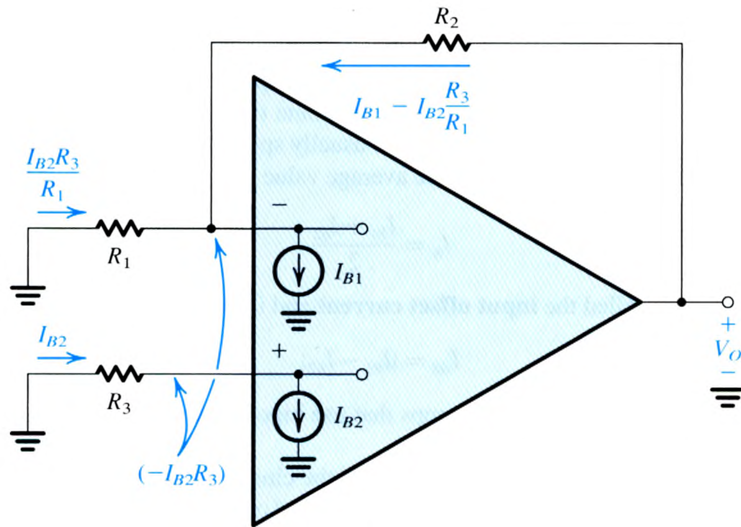


Figure 2.35 Reducing the effect of the input bias currents by introducing a resistor R_3.

method consists of introducing a resistance R_3 in series with the noninverting input lead, as shown in Fig. 2.35. From a signal point of view, R_3 has a negligible effect (ideally no effect). The appropriate value for R_3 can be determined by analyzing the circuit in Fig. 2.35, where analysis details are shown, and the output voltage is given by

$$V_O = -I_{B2}R_3 + R_2(I_{B1} - I_{B2}R_3/R_1) \tag{2.36}$$

Consider first the case $I_{B1} = I_{B2} = I_B$, which results in

$$V_O = I_B[R_2 - R_3(1 + R_2/R_1)]$$

Thus we can reduce V_O to zero by selecting R_3 such that

$$R_3 = \frac{R_2}{1 + R_2/R_1} = \frac{R_1 R_2}{R_1 + R_2} \tag{2.37}$$

That is, R_3 should be made equal to the parallel equivalent of R_1 and R_2.

Substituting Eq. (2.37) for R_3 into Eq. (2.36), and replacing $(I_{B1} - I_{B2})$ by I_{OS}, gives

$$V_O = I_{OS}R_2 \tag{2.38}$$

which is usually about an order of magnitude smaller than the value obtained without R_3 (Eq. 2.37). We conclude that to minimize the effect of the input bias currents, one should *place in the positive lead a resistance equal to the equivalant* dc *resistance seen by the inverting terminal.* We emphasize the word *dc* in the last statement; note that if the amplifier is ac-coupled, we should select $R_3 = R_2$, as shown in Fig. 2.36.

While we are on the subject of ac-coupled amplifiers, we should note that one must always provide a continuous dc path between each of the input terminals of the op amp and ground. This is the case no matter how small I_B is. For this reason the ac-coupled noninverting amplifier of Fig. 2.37 will *not* work without the resistance R_3 to ground. Unfortunately, including R_3 lowers considerably the input resistance of the closed-loop amplifier.

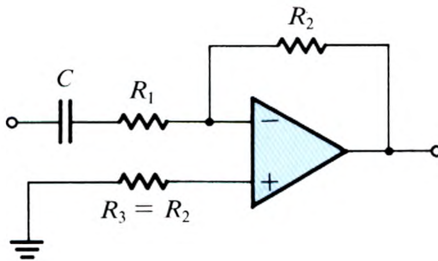


Figure 2.36 In an ac-coupled amplifier the dc resistance seen by the inverting terminal is R_2; hence R_3 is chosen equal to R_2.

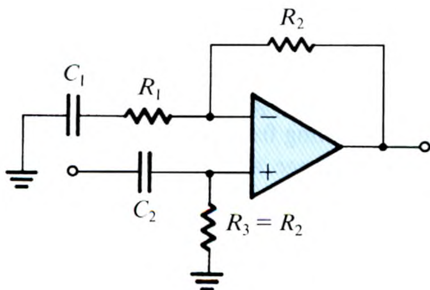


Figure 2.37 Illustrating the need for a continuous dc path for each of the op-amp input terminals. Specifically, note that the amplifier will *not* work without resistor R_3.

Video Example VE 2.3

An op amp is connected in a closed loop with gain of +10 using a feedback resistor of 1 MΩ.

(a) If the input bias current is 20 nA, what output voltage results with the input grounded?
(b) If the input offset voltage is ±2 mV and the input bias current as in (a), what is the largest possible output that can be observed with the input grounded?
(c) If bias-current compensation is used, what is the value of the required resistor? If the offset current is no more than one-tenth the bias current, what is the resulting output offset voltage (due to offset current alone)?
(d) With bias-current compensation as in (c) in place, what is the largest dc voltage at the output due to the combined effect of offset voltage and offset current?

VE 2.3

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: **2.104**

EXERCISE

2.24 Consider an inverting amplifier circuit designed using an op amp and two resistors, $R_1 = 10$ kΩ and $R_2 = 1$ MΩ. If the op amp is specified to have an input bias current of 100 nA and an input offset current of 10 nA, find the resulting output dc offset voltage and the value of a resistor R_3 placed in series with the positive input lead in order to minimize the output offset voltage. What is the new value of V_O?
Ans. 0.1 V; 9.9 kΩ ($\simeq$ 10 kΩ); 0.01 V

2.6.3 Effect of V_{OS} and I_{OS} on the Operation of the Inverting Integrator

In our discussion of the inverting integrator circuit in Section 2.5.2 we mentioned the susceptibility of this circuit to saturation in the presence of small dc voltages or currents. It behooves us, then, to consider the effect of the op-amp dc offsets on its operation. As you will see, these effects can be quite dramatic.

To see the effect of the input dc offset voltage V_{OS}, consider the integrator circuit in Fig. 2.38, where for simplicity we have short-circuited the input signal source. Analysis of the circuit is straightforward and is shown in Fig. 2.37. Assuming for simplicity that at time $t = 0$ the voltage across the capacitor is zero, the output voltage as a function of time is given by

$$v_O = V_{OS} + \frac{V_{OS}}{CR}t \tag{2.39}$$

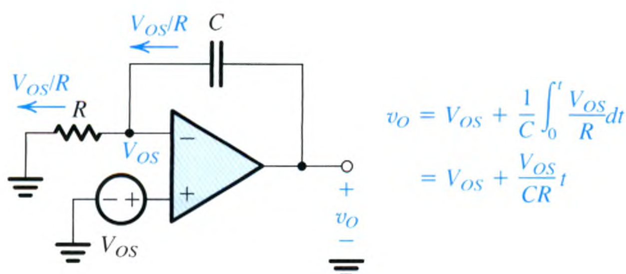


Figure 2.38 Determining the effect of the op-amp input offset voltage V_{OS} on the Miller integrator circuit. Note that since the output rises with time, the op amp eventually saturates.

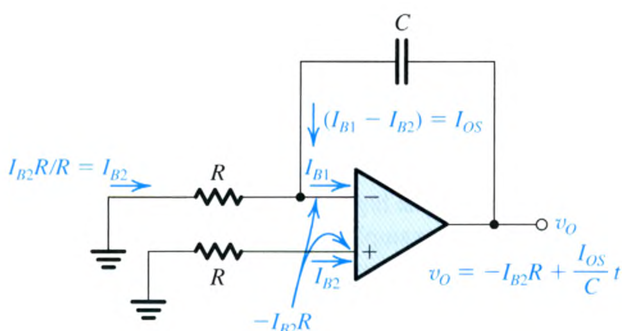


Figure 2.39 Effect of the op-amp input bias and offset currents on the performance of the Miller integrator circuit.

Thus v_O increases linearly with time until the op amp saturates—clearly an unacceptable situation! As should be expected, the dc input offset current I_{OS} produces a similar problem. Figure 2.39 illustrates the situation. Observe that we have added a resistance R in the op-amp positive-input lead in order to keep the input bias current I_B from flowing through C. Nevertheless, the offset current I_{OS} will flow through C and cause v_O to ramp linearly with time until the op amp saturates.

As mentioned in Section 2.5.2 the dc problem of the integrator circuit can be alleviated by connecting a resistor R_F across the integrator capacitor C, as shown in Fig. 2.26. Such a resistor provides a dc path through which the dc currents (V_{OS}/R) and I_{OS} can flow (assuming a resistance equal to $R \| R_F$ is connected in the positive op-amp lead), with the result that v_O will now have a dc component $[V_{OS}(1+R_F/R)+I_{OS}R_F]$ instead of rising linearly. To keep the dc offset at the output small, one would select a low value for R_F. Unfortunately, however, the lower the value of R_F, the less ideal the integrator circuit becomes.

EXERCISE

2.25 Consider a Miller integrator with a time constant of 1 ms and an input resistance of 10 kΩ. Let the op amp have $V_{OS} = 2$ mV and output saturation voltages of ±12 V. (a) Assuming that when the power supply is turned on the capacitor voltage is zero, how long does it take for the amplifier to saturate?

(b) Select the largest possible value for a feedback resistor R_F so that at least ±10 V of output signal swing remains available. What is the corner frequency of the resulting STC network?
Ans. (a) 6 s; (b) 10 MΩ, 0.16 Hz

2.7 Effect of Finite Open-Loop Gain and Bandwidth on Circuit Performance

2.7.1 Frequency Dependence of the Open-Loop Gain

The differential open-loop gain A of an op amp is not infinite; rather, it is finite and decreases with frequency. Figure 2.40 shows a plot for $|A|$, with the numbers typical of some commercially available general-purpose op amps (such as the popular 741-type op amp, available from many semiconductor manufacturers).

Note that although the gain is quite high at dc and low frequencies, it starts to fall off at a rather low frequency (10 Hz in our example). The uniform –20-dB/decade gain rolloff shown is typical of **internally compensated** op amps. These are units that have a network (usually a single capacitor) included within the same IC chip whose function is to cause the op-amp gain to have the single-time-constant (STC) low-pass response shown. This process of modifying the open-loop gain is termed **frequency compensation**, and its purpose is to ensure that op-amp circuits will be stable (as opposed to oscillatory). The subject of stability of op-amp circuits—or, more generally, of feedback amplifiers—will be studied in Chapter 11.

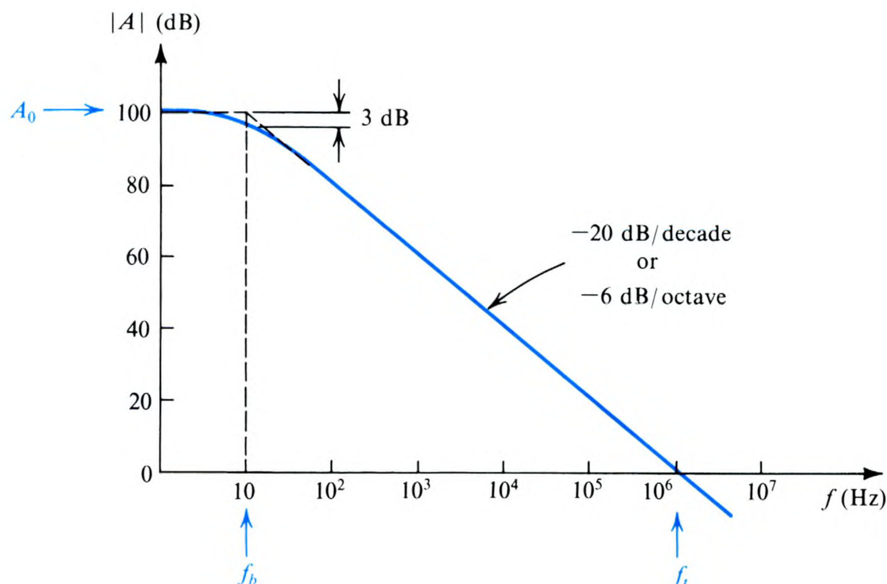


Figure 2.40 Open-loop gain of a typical general-purpose, internally compensated op amp.

By analogy to the response of low-pass STC circuits (see Section 1.6 and, for more detail, Appendix E), the gain $A(s)$ of an internally compensated op amp may be expressed as

$$A(s) = \frac{A_0}{1 + s/\omega_b} \tag{2.40}$$

which for physical frequencies, $s = j\omega$, becomes

$$A(j\omega) = \frac{A_0}{1 + j\omega/\omega_b} \tag{2.41}$$

where A_0 denotes the dc gain and ω_b is the 3-dB frequency (corner frequency or "break" frequency). For the example shown in Fig. 2.40, $A_0 = 10^5$ and $\omega_b = 2\pi \times 10$ rad/s. For frequencies $\omega \gg \omega_b$ (about 10 times and higher) Eq. (2.43) may be approximated by

$$A(j\omega) \simeq \frac{A_0\omega_b}{j\omega} \tag{2.42}$$

Thus,

$$|A(j\omega)| = \frac{A_0\omega_b}{\omega} \tag{2.43}$$

from which it can be seen that the gain $|A|$ reaches unity (0 dB) at a frequency denoted by ω_t and given by

$$\omega_t = A_0\omega_b \tag{2.44}$$

Substituting in Eq. (2.42) gives

$$A(j\omega) \simeq \frac{\omega_t}{j\omega} \tag{2.45}$$

The frequency $f_t = \omega_t/2\pi$ is usually specified on the data sheets of commercially available op amps and is known as the **unity-gain bandwidth**.[6] Also note that for $\omega \gg \omega_b$ the open-loop gain in Eq. (2.40) becomes

$$A(s) \simeq \frac{\omega_t}{s} \tag{2.46}$$

The gain magnitude can be obtained from Eq. (2.45) as

$$|A(j\omega)| \simeq \frac{\omega_t}{\omega} = \frac{f_t}{f} \tag{2.47}$$

Thus if f_t is known (10^6 Hz in our example), we can easily determine the magnitude of the op-amp gain at a given frequency f. Furthermore, observe that this relationship correlates with the Bode plot in Fig. 2.40. Specifically, for $f \gg f_b$, doubling f (an octave increase) results in halving the gain (a 6-dB reduction). Similarly, increasing f by a factor of 10 (a decade increase) results in reducing $|A|$ by a factor of 10 (20 dB).

As a matter of practical importance, we note that the production spread in the value of f_t between op-amp units of the same type is usually much smaller than that observed for A_0 and f_b. For this reason f_t is preferred as a specification parameter. Finally, it should be

[6]Since f_t is the product of the dc gain A_0 and the 3-dB bandwidth f_b (where $f_b = \omega_b/2\pi$), it is also known as the **gain–bandwidth product**. The reader is cautioned that in some amplifiers (those that do not have an STC response), the unity-gain frequency and the gain–bandwidth product are *not* equal.

mentioned that an op amp having this uniform –6-dB/octave (or equivalently –20-dB/decade) gain rolloff is said to have a **single-pole model**. Also, since this single pole *dominates* the amplifier frequency response, it is called a *dominant pole*. For more on poles (and zeros), the reader may wish to consult Appendix F.

EXERCISE

2.26 An internally compensated op amp is specified to have an open-loop dc gain of 106 dB and a unity-gain bandwidth of 3 MHz. Find f_b and the open-loop gain (in dB) at f_b, 300 Hz, 3 kHz, 12 kHz, and 60 kHz.
Ans. 15 Hz; 103 dB; 80 dB; 60 dB; 48 dB; 34 dB

2.7.2 Frequency Response of Closed-Loop Amplifiers

We next consider the effect of limited op-amp gain and bandwidth on the closed-loop transfer functions of the two basic configurations: the inverting circuit of Fig. 2.5 and the noninverting circuit of Fig. 2.12. The closed-loop gain of the inverting amplifier, assuming a finite op-amp open-loop gain A, was derived in Section 2.2 and given in Eq. (2.5), which we repeat here.

$$\frac{V_o}{V_i} = \frac{-R_2/R_1}{1+(1+R_2/R_1)/A} \tag{2.48}$$

Substituting for A from Eq. (2.40) and using Eq. (2.44) gives

$$\frac{V_o(s)}{V_i(s)} = \frac{-R_2/R_1}{1+\dfrac{1}{A_0}\left(1+\dfrac{R_2}{R_1}\right)+\dfrac{s}{\omega_t/(1+R_2/R_1)}} \tag{2.49}$$

For $A_0 \gg 1+R_2/R_1$, which is usually the case,

$$\frac{V_o(s)}{V_i(s)} \simeq \frac{-R_2/R_1}{1+\dfrac{s}{\omega_t/(1+R_2/R_1)}} \tag{2.50}$$

which is of the same form as that for a low-pass STC network (see Table 1.2). Thus the inverting amplifier has an STC low-pass response with a dc gain of magnitude equal to R_2/R_1. The closed-loop gain rolls off at a uniform –20-dB/decade slope with a corner frequency (3-dB frequency) given by

$$\omega_{3\text{dB}} = \frac{\omega_t}{1+R_2/R_1} \tag{2.51}$$

Similarly, analysis of the noninverting amplifier of Fig. 2.12, assuming a finite open-loop gain A, yields the closed-loop gain in Eq. (2.11), repeated here.

$$\frac{V_o}{V_i} = \frac{1 + R_2/R_1}{1 + (1 + R_2/R_1)/A} \tag{2.52}$$

Substituting for A from Eq. (2.40) and making the approximation $A_0 \gg 1 + R_2/R_1$ results in

$$\frac{V_o(s)}{V_i(s)} \simeq \frac{1 + R_2/R_1}{1 + \dfrac{s}{\omega_t/(1 + R_2/R_1)}} \tag{2.53}$$

Thus the noninverting amplifier has an STC low-pass response with a dc gain of $(1 + R_2/R_1)$ and a 3-dB frequency given also by Eq. (2.51).

Example 2.7

Consider an op amp with $f_t = 1$ MHz. Find the 3-dB frequency of closed-loop amplifiers with nominal gains of +1000, +100, +10, +1, −1, −10, −100, and −1000. Sketch the magnitude frequency response for the amplifiers with closed-loop gains of +10 and −10.

Solution

We use Eq. (2.51) to obtain the results given in the following table.

Closed-Loop Gain	R_2/R_1	$f_{3\text{ dB}} = f_t/(1 + R_2/R_1)$
+1000	999	1 kHz
+100	99	10 kHz
+10	9	100 kHz
+1	0	1 MHz
−1	1	0.5 MHz
−10	10	90.9 kHz
−100	100	9.9 kHz
−1000	1000	$\simeq$ 1 kHz

Figure 2.41 shows the frequency response for the amplifier whose nominal dc gain is +10 (20 dB), and Fig. 2.42 shows the frequency response for the –10 (also 20 dB) case. An interesting observation follows from the table above: The unity-gain inverting amplifier has a 3-dB frequency of $f_t/2$ as compared to f_t for the unity-gain noninverting amplifier (the unity-gain voltage follower).

Example 2.7 *continued*

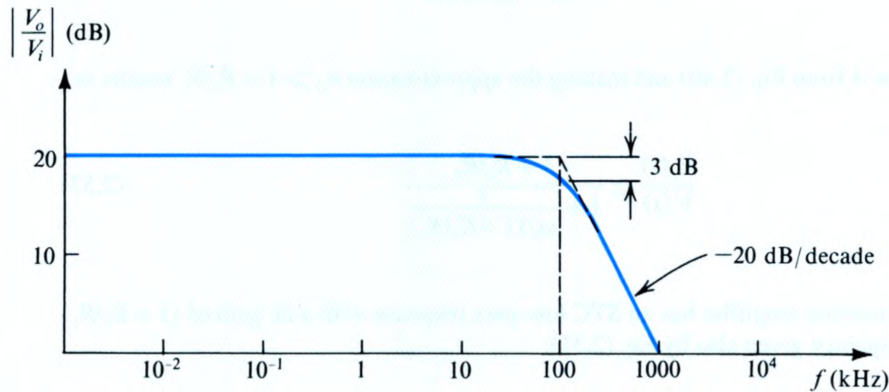


Figure 2.41 Frequency response of an amplifier with a nominal gain of +10 V/V.

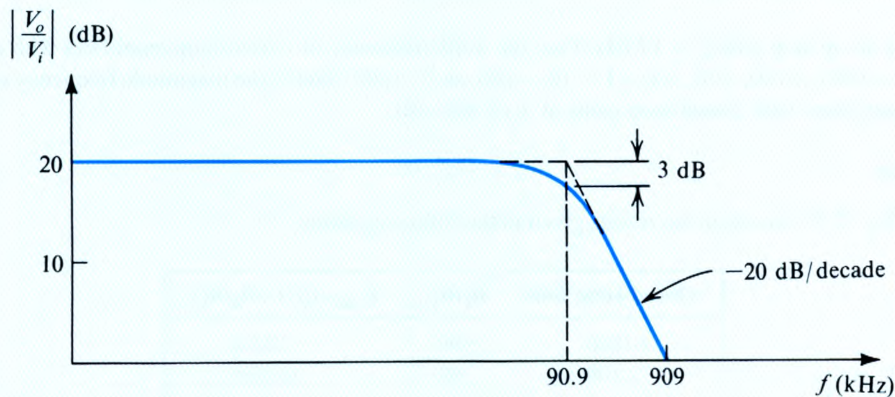


Figure 2.42 Frequency response of an amplifier with a nominal gain of −10 V/V.

The table in Example 2.7 above clearly illustrates the trade-off between gain and bandwidth: For a given op amp, the lower the closed-loop gain required, the wider the bandwidth achieved. Indeed, the noninverting configuration exhibits a constant **gain–bandwidth product** equal to f_t of the op amp. An interpretation of these results in terms of feedback theory will be given in Chapter 11.

Video Example VE 2.4

Consider a noninverting amplifier with a gain of 10 using an internally compensated op amp with f_t = 4 MHz. What is the 3-dB frequency of the amplifier? If the input is a 10 mV-step, find the 10% to 90% rise time of the output voltage. (*Note*: The step response of STC low-pass networks is discussed in Appendix E. Specifically, note that the 10% to 90% rise time of a low-pass STC circuit with a time constant τ is 2.2τ.)

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

VE 2.4

Related end-of-chapter problem: **2.115**

EXERCISES

2.27 An internally compensated op amp has a dc open-loop gain of 10^6 V/V and an open-loop gain of 40 dB at 10 kHz. Estimate its 3-dB frequency, its unity-gain frequency, its gain–bandwidth product, and its expected gain at 1 kHz.
Ans. 1 Hz; 1 MHz; 1 MHz; 60 dB

2.28 An op amp having a 106-dB gain at dc and a single-pole frequency response with $f_t = 2$ MHz is used to design a noninverting amplifier with nominal dc gain of 100. Find the 3-dB frequency of the closed-loop gain.
Ans. 20 kHz

2.8 Large-Signal Operation of Op Amps

In this section, we study the limitations on the performance of op-amp circuits when large output signals are present.

2.8.1 Output Voltage Saturation

Similar to all other amplifiers, op amps operate linearly over a limited range of output voltages. Specifically, the op-amp output saturates in the manner shown in Fig. 1.14 with L_+ and L_- related to the positive and negative power supply voltages, respectively. Thus, an op amp that is operating from ±15-V supplies may saturate when the output voltage reaches about +13 V in the positive direction and –13 V in the negative direction. For this particular op amp the **rated output voltage** is said to be ±13 V. To avoid clipping off the peaks of

the output waveform, and the resulting waveform distortion, the input signal must be kept correspondingly small. When operating from low supply voltages, such as those provided by single-cell batteries, it is naturally important for $L_{\pm}$ to closely approach the supply voltages, thus allowing for the largest possible signal amplitude. Op amps for which $L_{\pm}$ equal the supply voltages are common for such applications and are said to offer **rail-to-rail** output.

2.8.2 Output Current Limits

Another limitation on the operation of op amps is that their output current is limited to a specified maximum. For instance, the popular 741 op amp is specified to have a maximum output current of ±20 mA. Thus, in designing closed-loop circuits utilizing the 741, the designer has to ensure that under no condition will the op amp be required to supply an output current, in either direction, exceeding 20 mA. This, of course, has to include both the current in the feedback circuit as well as the current supplied to a load resistor. If the circuit requires a larger current, the op-amp output voltage will saturate at the level corresponding to the maximum allowed output current.

Example 2.8

Consider the noninverting amplifier circuit shown in Fig. 2.43. As shown, the circuit is designed for a nominal gain $(1 + R_2/R_1) = 10$ V/V. It is fed with a low-frequency sine-wave signal of peak voltage V_p and is connected to a load resistor R_L. The op amp is specified to have output saturation voltages of ±13 V and output current limits of ±20 mA.

(a) For $V_p = 1$ V and $R_L = 1$ kΩ, specify the signal resulting at the output of the amplifier.
(b) For $V_p = 1.5$ V and $R_L = 1$ kΩ, specify the signal resulting at the output of the amplifier.
(c) For $R_L = 1$ kΩ, what is the maximum value of V_p for which an undistorted sine-wave output is obtained?
(d) For $V_p = 1$ V, what is the lowest value of R_L for which an undistorted sine-wave output is obtained?

Solution

(a) For $V_p = 1$ V and $R_L = 1$ kΩ, the output will be a sine wave with peak value of 10 V. This is lower than output saturation levels of ±13 V, and thus the amplifier is not limited that way. Also, when the output is at its peak (10 V), the current in the load will be 10 V/1 kΩ = 10 mA, and the current in the feedback network will be 10 V/(9 + 1) kΩ = 1 mA, for a total op-amp output current of 11 mA, well under its limit of 20 mA.

(b) Now if V_p is increased to 1.5 V, ideally the output would be a sine wave of 15-V peak. The op amp, however, will saturate at ±13 V, thus clipping the sine-wave output at these levels. Let's next check on the op-amp output current: At 13-V output and $R_L = 1$ kΩ, $i_L = 13$ mA and $i_F = 1.3$ mA; thus $i_O = 14.3$ mA, again under the 20-mA limit. Thus the output will be a sine wave with its peaks clipped off at ±13 V, as shown in Fig. 2.43(b).

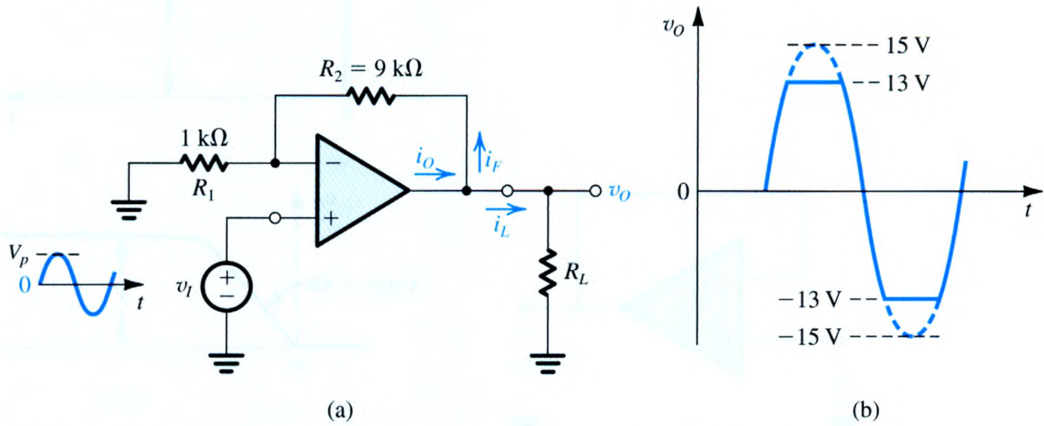


Figure 2.43 **(a)** A noninverting amplifier with a nominal gain of 10 V/V designed using an op amp that saturates at ±13-V output voltage and has ±20-mA output current limits. **(b)** When the input sine wave has a peak of 1.5 V, the output is clipped off at ±13 V.

(c) For $R_L = 1\ \text{k}\Omega$, the maximum value of V_p for undistorted sine-wave output is 1.3 V. The output will be a 13-V peak sine wave, and the op-amp output current at the peaks will be 14.3 mA.

(d) For $V_p = 1$ V and R_L reduced, the lowest value possible for R_L while the output is remaining an undistorted sine wave of 10-V peak can be found from

$$i_{O\max} = 20\ \text{mA} = \frac{10\ \text{V}}{R_{L\min}} + \frac{10\ \text{V}}{9\ \text{k}\Omega + 1\ \text{k}\Omega}$$

which results in

$$R_{L\min} = 526\ \Omega$$

2.8.3 Slew Rate

Another phenomenon that can cause nonlinear distortion when large output signals are present is slew-rate limiting. The name refers to the fact that there is a specific *maximum rate of change* possible at the output of a real op amp. This maximum is known as the **slew rate** (SR) of the op amp and is defined as

$$\text{SR} = \left.\frac{dv_O}{dt}\right|_{\max} \tag{2.54}$$

and is usually specified on the op-amp data sheet in units of V/μs. It follows that if the input signal applied to an op-amp circuit is such that it demands an output response that is faster than the specified value of SR, the op amp will not comply. Rather, its output will change at the maximum possible rate, which is equal to its SR. As an example, consider an op amp connected in the unity-gain voltage-follower configuration shown in Fig. 2.44(a), and let the

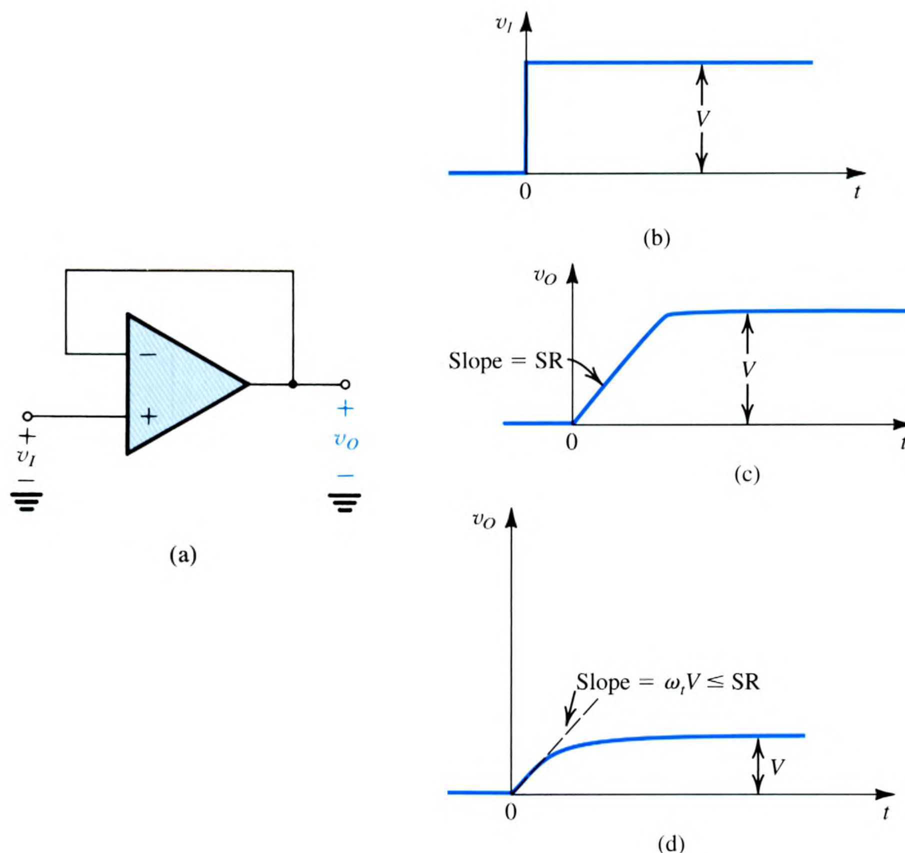


Figure 2.44 **(a)** Unity-gain follower. **(b)** Input step waveform. **(c)** Linearly rising output waveform obtained when the amplifier is slew-rate limited. **(d)** Exponentially rising output waveform obtained when V is sufficiently small so that the initial slope ($\omega_t V$) is smaller than or equal to SR.

input signal be the step voltage shown in Fig. 2.44(b). The output of the op amp will not be able to rise instantaneously to the ideal value V; rather, the output will be the linear ramp of slope equal to SR, shown in Fig. 2.44(c). The amplifier is then said to be **slewing**, and its output is **slew-rate limited**.

In order to understand the origin of the slew-rate phenomenon, we need to know about the internal circuit of the op amp, which we will study in Chapter 13. For now, it is enough to be aware of the phenomenon and that it is distinct from the finite op-amp bandwidth that limits the frequency response of the closed-loop amplifiers, studied in the previous section. The limited bandwidth is a linear phenomenon and does not result in a change in the shape of an input sinusoid; that is, it does not lead to nonlinear distortion. The slew-rate limitation, on the other hand, can cause nonlinear distortion to an input sinusoidal signal when its frequency and amplitude are such that the corresponding ideal output would require v_O to change at a rate greater than SR. This is the origin of another related op-amp specification, its full-power bandwidth, to be explained later.

Before leaving the example in Fig. 2.44, however, we should point out that if the step input voltage V is sufficiently small, the output can be the exponentially rising ramp shown

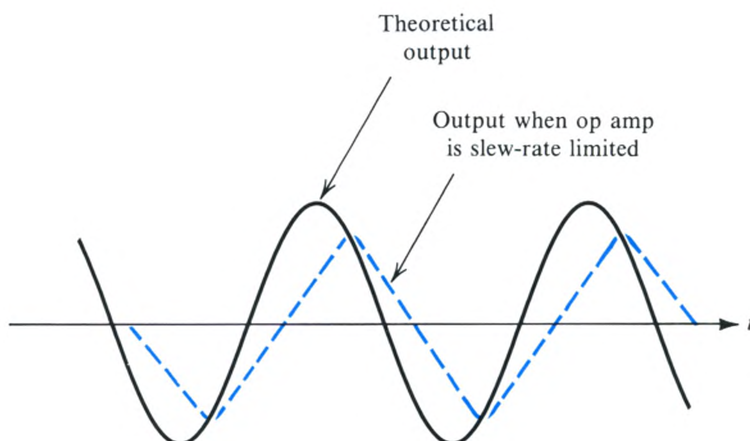


Figure 2.45 Effect of slew-rate limiting on output sinusoidal waveforms.

in Fig. 2.44(d). Such an output would be expected from the follower if the only limitation on its dynamic performance were the finite op-amp bandwidth. Specifically, the transfer function of the follower can be found by substituting $R_1 = \infty$ and $R_2 = 0$ in Eq. (2.53) to obtain

$$\frac{V_o}{V_i} = \frac{1}{1 + s/\omega_t} \tag{2.55}$$

which is a low-pass STC response with a time constant $1/\omega_t$. Its step response would therefore be (see Appendix E)

$$v_O(t) = V(1 - e^{-\omega_t t}) \tag{2.56}$$

The initial slope of this exponentially rising function is $(\omega_t V)$. Thus, as long as V is sufficiently small so that $\omega_t V \leq \text{SR}$, the output will be as in Fig. 2.44(d).

EXERCISE

2.29 An op amp that has a slew rate of 1 V/μs and a unity-gain bandwidth f_t of 1 MHz is connected in the unity-gain follower configuration. Find the largest possible input voltage step for which the output waveform will still be given by the exponential ramp of Eq. (2.56). For this input voltage, what is the 10% to 90% rise time of the output waveform? If an input step 10 times as large is applied, find the 10% to 90% rise time of the output waveform.

Ans. 0.16 V; 0.35 μs; 1.28 μs

Example 2.9

Consider the unity-gain follower with a sine-wave input given by

$$v_I = \hat{V}_i \sin\omega t$$

An op amp having a slew rate SR and maximum output swing V_{omax}, but otherwise ideal, is used. What is the maximum frequency, f_M, for which a full-swing sinusoid will not cause distortion due to slew-rate limiting?

Solution

In a unity-gain follower, the input and output voltage waveforms are the same. The rate of change of a full-swing sinusoid, $v_I(t) = V_{omax} \sin\omega t$, is

$$\frac{dv_I}{dt} = \omega V_{omax} \cos\omega t$$

with a maximum value of ωV_{omax} at the zero crossings of the sinusoid. This value will equal the op-amp slew rate at a frequency

$$\omega_M = \frac{\text{SR}}{V_{omax}} \Rightarrow f_M = \frac{\text{SR}}{2\pi V_{omax}} \tag{2.57}$$

Beyond this frequency, the output waveform will be distorted as illustrated in Fig. 2.45. Op-amp data sheets usually specify f_M, called the **full-power bandwidth**.

At smaller output amplitudes, higher frequencies can be used without slew-rate limiting. In general, at a frequency ω, the maximum amplitude of the undistorted output sinusoid is

$$V_o = V_{omax}\left(\frac{\omega_M}{\omega}\right) \tag{2.58}$$

EXERCISE

2.30 An op amp has a rated output voltage of ±5 V and a slew rate of 10 V/μs. What is its full-power bandwidth? If an input sinusoid with frequency $f = 5f_M$ is applied to a unity-gain follower constructed using this op amp, what is the maximum possible amplitude that can be accommodated at the output without incurring SR distortion?

Ans. 318 kHz; 1 V (peak)

Summary

- The IC op amp is a versatile circuit building block. It is easy to apply, and the performance of op-amp circuits closely matches theoretical predictions.
- The op-amp terminals are the inverting input terminal (1), the noninverting input terminal (2), the output terminal (3), the positive-supply terminal (4) to be connected to the positive power supply (V_{CC}), and the negative-supply terminal (5) to be connected to the negative supply ($-V_{EE}$).
- The ideal op amp responds only to the difference input signal, that is, $(v_2 - v_1)$. It provides at the output terminal 3 a signal $A(v_2 - v_1)$, where A, the open-loop gain, is very large (e.g., 10^4 to 10^6) and ideally infinite. It has infinite input resistance and zero output resistance. (See Table 2.1.)
- Negative feedback is applied to an op amp by connecting a passive component between its output terminal and its inverting (negative) input terminal. Negative feedback causes the voltage between the two input terminals to become very small and ideally zero. Correspondingly, a virtual short circuit is said to exist between the two input terminals. If the positive input terminal is connected to ground, a virtual ground appears on the negative input terminal.
- The two most important assumptions in the analysis of op-amp circuits, presuming negative feedback and ideal op amps, are as follows: the two input terminals of the op amp are at the same voltage, and zero current flows into the op-amp input terminals.
- With negative feedback applied and the loop closed, the closed-loop gain is almost entirely determined by external components: For the inverting configuration, $V_o/V_i = -R_2/R_1$; and for the noninverting configuration, $V_o/V_i = 1 + R_2/R_1$.
- The noninverting closed-loop configuration features a very high input resistance. A special case is the unity-gain follower, frequently employed as a buffer amplifier to connect a high-resistance source to a low-resistance load.
- The difference amplifier of Fig. 2.16 is designed with $R_4/R_3 = R_2/R_1$, resulting in $v_O = (R_2/R_1)(v_{I2} - v_{I1})$.
- The instrumentation amplifier of Fig. 2.20(b) provides $v_O = (1 + R_2/R_1)(R_4/R_3)\ (v_{I2} - v_{I1})$. It is usually designed with $R_3 = R_4$, and R_1 and R_2 selected to provide the required gain. If an adjustable gain is needed, part of R_1 can be made variable.
- The inverting Miller integrator of Fig. 2.24(a) is frequently used in analog signal-processing functions such as filters (Chapter 14).
- The input offset voltage, V_{OS}, is the magnitude of dc voltage that when applied between the op-amp input terminals, with appropriate polarity, reduces the dc offset voltage at the output to zero.
- The effect of V_{OS} on performance can be evaluated by including in the analysis a dc source V_{OS} in series with the op-amp positive input lead. For both the inverting and the noninverting configurations, V_{OS} results in a dc offset voltage at the output of $V_{OS}(1 + R_2/R_1)$.
- Capacitively coupling an op amp reduces the dc offset voltage at the output considerably.
- The average of the two dc currents, I_{B1} and I_{B2}, that flow into the input terminals of the op amp, is called the input bias current, I_B. In a closed-loop amplifier, I_B gives rise to a dc offset voltage at the output of magnitude $I_B R_2$. This voltage can be reduced to $I_{OS}R_2$ by connecting a resistance in series with the positive input terminal equal to the total dc resistance seen by the negative input terminal, where I_{OS} is the input offset current; that is, $I_{OS} = |I_{B1} - I_{B2}|$.
- Connecting a large resistance in parallel with the capacitor of an op-amp inverting integrator prevents op-amp saturation (due to the effect of V_{OS} and I_B).
- For most internally compensated op amps, the open-loop gain falls off with frequency at a rate of −20 dB/decade, reaching unity at a frequency f_t (the unity-gain bandwidth). Frequency f_t is also known as the gain–bandwidth product of the op amp: $f_t = A_0 f_b$, where A_0 is the dc gain, and f_b is the 3-dB frequency of the open-loop gain. At any frequency $f \gg f_b$, the op-amp gain $|A| \simeq f_t/f$.
- For both the inverting and the noninverting closed-loop configurations, the 3-dB frequency is equal to $f_t/(1 + R_2/R_1)$.
- The maximum rate at which the op-amp output voltage can change is called the slew rate. The slew rate, SR, is usually specified in V/μs. Op-amp slewing can result in nonlinear distortion of output signal waveforms.

Computer Simulation Problems

SIM Problems identified by this icon are intended to demonstrate the value of using SPICE simulation to verify hand analysis and design, and to investigate important issues such as allowable signal swing and amplifier nonlinear distortion. Instructions to assist in setting up simulations for all the indicated problems can be found in the corresponding files on the website. Note that if a particular parameter value is not specified in the problem statement, you are to make a reasonable assumption.

Section 2.1: The Ideal Op Amp

2.1 What is the minimum number of pins required for a so-called dual-op-amp IC package, one containing two op amps? What is the number of pins required for a so-called quad-op-amp package, one containing four op amps?

2.2 The circuit of Fig. P2.2 uses an op amp that is ideal except for having a finite gain A. Measurements indicate $v_O = 2.5$ V when $v_I = 1.0$ V. What is the op-amp gain A?

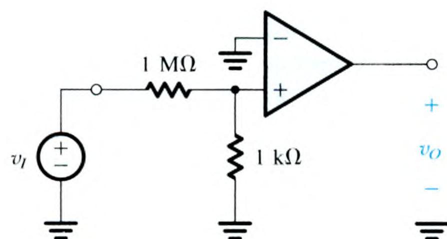


Figure P2.2

2.3 Measurement of a circuit incorporating what is thought to be an ideal op amp shows the voltage at the op-amp output to be −3.500 V and that at the negative input to be −1.000 V. For the amplifier to be ideal, what would you expect the voltage at the positive input to be? If the measured voltage at the positive input is −1.002 V, what is likely to be the actual gain of the amplifier?

2.4 A set of experiments is run on an op amp that is ideal except for having a finite gain A. The results are tabulated below. Are the results consistent? If not, are they reasonable, in view of the possibility of experimental error? What do they show the gain to be? Using this value, predict values of the measurements that were accidentally omitted (the blank entries).

Experiment #	v_1	v_2	v_O
1	0.00	0.00	0.00
2	−1.00	−1.00	0.00
3		−1.00	−1.00
4	1.00	1.02	4.01
5	2.01	2.00	−1.99
6	1.99	2.00	2.00
7	5.10		−5.10

***2.5** Refer to Exercise 2.3. This problem explores an alternative internal structure for the op amp. In particular, we wish to model the internal structure of a particular op amp using two transconductance amplifiers and one transresistance amplifier. Suggest an appropriate topology. For equal transconductances G_m and a transresistance R_m, find an expression for the open-loop gain A. For $G_m = 100$ mA/V and $R_m = 1 \times 10^5\ \Omega$, what value of A results?

2.6 Two amplifier inputs are expressed in each row of the table below either in terms of their node voltages with respect to ground, v_1 and v_2, or in terms of their differential and common-mode voltages, v_{Id} and v_{Icm}. Complete the table.

v_1	v_2	v_{Id}	v_{Icm}
0.90	0.90		
		0.50	1.00
−1.20			−1.10
−0.15	0.05		
		−0.20	0.00

2.7 The two wires leading from the output terminals of a transducer pick up an interference signal that is a 60-Hz, 2-V sinusoid. The output signal of the transducer is sinusoidal of 5-mV amplitude and 1000-Hz frequency. Give expressions for v_{cm}, v_d, and the total signal between each wire and the system ground.

2.8 Nonideal (i.e., real) operational amplifiers respond to both the differential and common-mode components of their input signals (refer to Fig. 2.4 for signal representation). Thus

the output voltage of the op amp can be expressed as

$$v_O = A_d v_{Id} + A_{cm} v_{Icm}$$

where A_d is the differential gain (referred to simply as A in the text) and A_{cm} is the common-mode gain (assumed to be zero in the text). The op amp's effectiveness in rejecting common-mode signals is measured by its CMRR, defined as

$$\text{CMRR} = 20\log\left|\frac{A_d}{A_{cm}}\right|$$

Consider an op amp whose internal structure is of the type shown in Fig. E2.3 except for a mismatch ΔG_m between the transconductances of the two channels; that is,

$$G_{m1} = G_m - \tfrac{1}{2}\Delta G_m$$

$$G_{m2} = G_m + \tfrac{1}{2}\Delta G_m$$

Find expressions for A_d, A_{cm}, and CMRR. What is the maximum permitted percentage mismatch between the two G_m values if a minimum CMRR of 60 dB is required?

Section 2.2: The Inverting Configuration

2.9 Assuming ideal op amps, find the voltage gain v_o/v_i and input resistance R_{in} of each of the circuits in Fig. P2.9.

2.10 A particular inverting circuit uses an ideal op amp and two 10-kΩ resistors. What closed-loop gain would you expect? If a dc voltage of +1.00 V is applied at the input, what outputs result? If the 10-kΩ resistors are said to be "1% resistors," having values somewhere in the range (1 ± 0.01) times the nominal value, what range of outputs would you expect to actually measure for an input of precisely 1.00 V?

2.11 You are provided with an ideal op amp and three 10-kΩ resistors. Using series and parallel resistor combinations, how many different inverting-amplifier circuit topologies are possible? What is the largest (noninfinite) available voltage gain magnitude? What is the smallest (nonzero) available gain magnitude? What are the input resistances in these two cases?

SIM 2.12 For ideal op amps operating with the following feedback networks in the inverting configuration, what closed-loop gain results?

(a) $R_1 = 10\text{ k}\Omega, R_2 = 20\text{ k}\Omega$
(b) $R_1 = 10\text{ k}\Omega, R_2 = 100\text{ k}\Omega$
(c) $R_1 = 10\text{ k}\Omega, R_2 = 5\text{ k}\Omega$
(d) $R_1 = 100\text{ k}\Omega, R_2 = 5\text{ M}\Omega$
(e) $R_1 = 100\text{ k}\Omega, R_2 = 0.5\text{ M}\Omega$

D 2.13 Given an ideal op amp, what are the values of the resistors R_1 and R_2 to be used to design amplifiers with the closed-loop gains listed below? In your designs, use at least one 10-kΩ resistor and another equal or larger resistor.

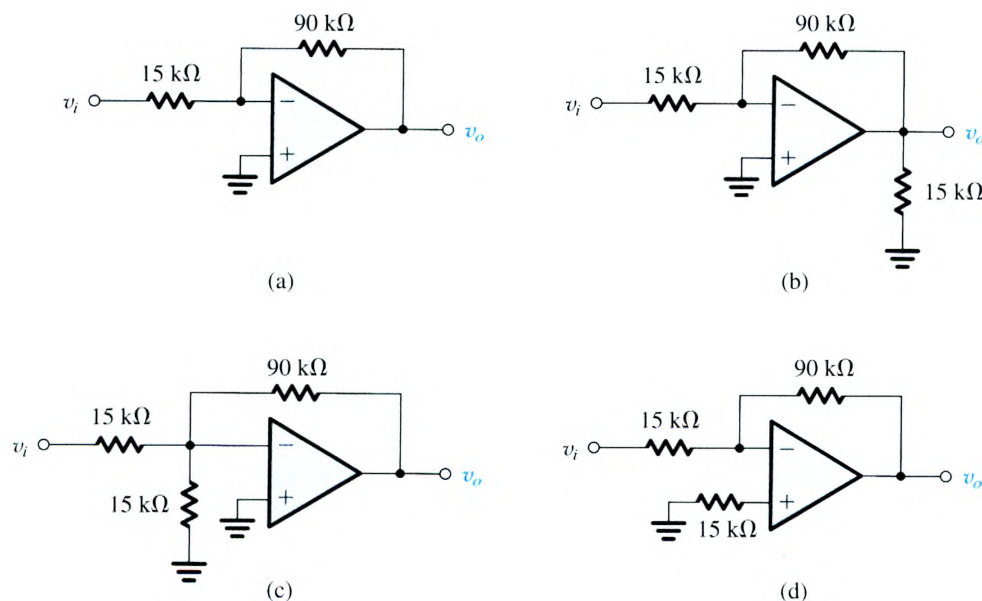


Figure P2.9

(a) −1 V/V
(b) −10 V/V
(c) −20 V/V
(d) −0.5 V/V

D 2.14 Design an inverting op-amp circuit for which the gain is −5 V/V and the total resistance used is 6 kΩ.

D 2.15 Using the circuit of Fig. 2.5 and assuming an ideal op amp, design an inverting amplifier with a gain of 26 dB having the largest possible input resistance under the constraint of having to use resistors no larger than 100 kΩ. What is the input resistance of your design?

VE 2.1 **2.16** An ideal op amp is connected as shown in Fig. 2.5 with $R_1 = 2$ kΩ and $R_2 = 8$ kΩ. A symmetrical square-wave signal with levels of 0 V and 1 V is applied at the input. Sketch and clearly label the waveform of the resulting output voltage. What is its average value? What is its highest value? What is its lowest value? Also sketch the current waveform through R_1, labeling its maximum, minimum, and average.

2.17 For the circuit in Fig. P2.17, assuming an ideal op amp, find the currents through all branches and the voltages at all nodes. Find the power dissipated in each resistor, and the total power dissipated in all resistors. How much power is delivered by the −1-V input source? Where does the rest of the power come from?

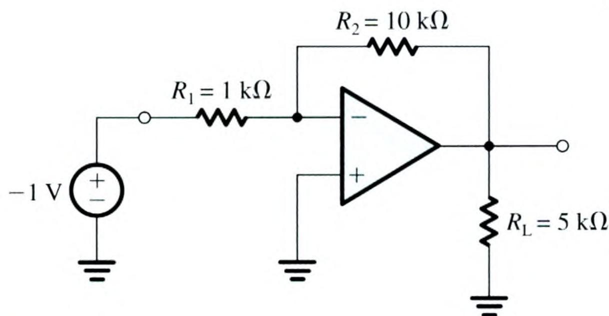


Figure P2.17

2.18 In the circuit of Fig. P2.17, what is the current flowing through the output of the ideal op amp? If we wish to use a practical op amp that has a specified maximum output current of 2.5 mA, how may the values of the resistors R_1 and R_2 be changed so that the gain remains the same?

2.19 An inverting op-amp circuit is fabricated with the resistors R_1 and R_2 having x% tolerance (i.e., the value of each resistance can deviate from the nominal value by as much as $\pm x$%). What is the tolerance on the realized closed-loop gain? Assume the op amp to be ideal. If the nominal closed-loop gain is −100 V/V and $x = 1$, what is the range of gain values expected from such a circuit?

2.20 An ideal op amp with 5-kΩ and 10-kΩ resistors is used to create a +2.5-V supply from a −5-V reference. Sketch the circuit. What are the voltages at the ends of the 5-kΩ resistor? If these resistors are so-called 1% resistors, whose actual values are the range bounded by the nominal value ±1%, what are the limits of the output voltage produced? If the −5-V supply can also vary by ±1%, what is the range of the output voltages that might be found?

D 2.21 An inverting op-amp circuit for which the required gain is −50 V/V uses an op amp whose open-loop gain is only 200 V/V. If the larger resistor used is 100 kΩ, to what must the smaller be adjusted? With what resistor must a 2-kΩ resistor connected to the input be shunted to achieve this goal? (Note that a resistor R_a is said to be shunted by resistor R_b when R_b is placed in parallel with R_a.)

D 2.22 (a) Design an inverting amplifier with a closed-loop gain of −200 V/V and an input resistance of 1 kΩ.
(b) If the op amp is known to have an open-loop gain of 5000 V/V, what do you expect the closed-loop gain of your circuit to be (assuming the resistors have precise values)?
(c) Give the value of a resistor you can place in parallel (shunt) with R_1 to restore the closed-loop gain to its nominal value. Use the closest standard 1% resistor value (see Appendix J).
(d) Alternatively, give the closest standard resistor value that may be connected in series with R_2 to restore the closed-loop gain to its ideal value.

2.23 An op amp with an open-loop gain of 5000 V/V is used in the inverting configuration so that its output voltage ranges from −10 V to +10 V. What is the maximum voltage by which the "virtual ground node" departs from its ideal value?

2.24 The circuit in Fig. P2.24 is frequently used to provide an output voltage v_o proportional to an input signal current i_i.

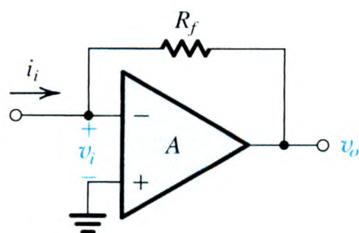


Figure P2.24

Derive expressions for the transresistance $R_m \equiv v_o/i_i$ and the input resistance $R_i \equiv v_i/i_i$ for the following cases:

(a) A is infinite.
(b) A is finite.

2.25 Show that for the inverting amplifier if the op-amp gain is A, the input resistance is given by

$$R_{in} = R_1 + \frac{R_2}{A+1}$$

***2.26** Figure P2.26 shows an op amp that is ideal except for having a finite open-loop gain and is used to realize an inverting amplifier whose gain has a nominal magnitude $G = R_2/R_1$. To compensate for the gain reduction due to the finite A, a resistor R_c is shunted across R_1. Show that perfect compensation is achieved when R_c is selected according to

$$\frac{R_c}{R_1} = \frac{A-G}{1+G}$$

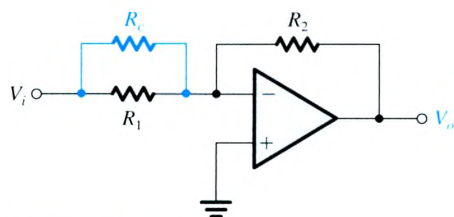


Figure P2.26

D*2.27 (a) Use Eq. (2.5) to obtain the amplifier open-loop gain A required to realize a specified closed-loop gain ($G_{nominal} = -R_2/R_1$) within a specified gain error ϵ,

$$\epsilon \equiv \left| \frac{G - G_{nominal}}{G_{nominal}} \right|$$

(b) Design an inverting amplifer for a nominal closed-loop gain of −30, an input resistance of 1 kΩ, and a gain error of ≤5%. Specify R_1, R_2, and the minimum A required.

***2.28** (a) Use Eq. (2.5) to show that a reduction ΔA in the op-amp gain A gives rise to a reduction $\Delta|G|$ in the magnitude of the closed-loop gain G with $\Delta|G|$ and ΔA related by

$$\frac{\Delta|G|/|G|}{\Delta A/A} \simeq \frac{1 + R_2/R_1}{A}$$

Assume that $\left(1 + \frac{R_2}{R_1}\right) \ll A$ and $\frac{\Delta A}{A} \ll 1$.

(b) If in a closed-loop amplifier with a nominal gain (i.e., R_2/R_1) of 100, A decreases by 10%, what is the minimum nominal A required to limit the percentage change in $|G|$ to 0.1%?

2.29 An inverting amplifier is accidentally connected with positive feedback as illustrated in Fig. P2.29. Assuming the op amp has a finite gain A, find an expression for the closed-loop gain, v_O/v_I. Show that if $A = 1 + (R_2/R_1)$, the closed-loop gain becomes infinite, indicating instability. Real op-amp gain varies with input frequency so that instability is practically inevitable when positive feedback is used.

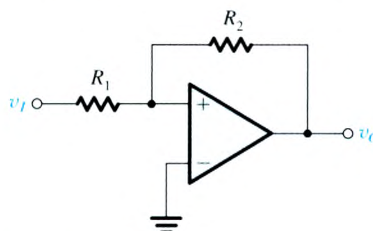


Figure P2.29

2.30 Consider the circuit in Fig. 2.8 with $R_1 = R_2 = R_4 =$ 2 MΩ, and assume the op amp to be ideal. Find values for R_3 to obtain the following gains:

(a) −50 V/V
(b) −20 V/V
(c) −5 V/V

D 2.31 An inverting op-amp circuit using an ideal op amp must be designed to have a gain of −200 V/V using resistors no larger than 1 MΩ.

(a) For the simple two-resistor circuit, what input resistance would result?
(b) If the circuit in Fig. 2.8 is used with three resistors of maximum value, what input resistance results? What is the value of the smallest resistor needed?

2.32 The inverting circuit with the T network in the feedback is redrawn in Fig. P2.32 in a way that emphasizes the

observation that R_2 and R_3 in effect are in parallel (because the ideal op amp forces a virtual ground at the inverting input terminal). Use this observation to derive an expression for the gain (v_O/v_I) by first finding (v_X/v_I) and (v_O/v_X). For the latter use the voltage-divider rule applied to R_4 and $(R_2 \| R_3)$.

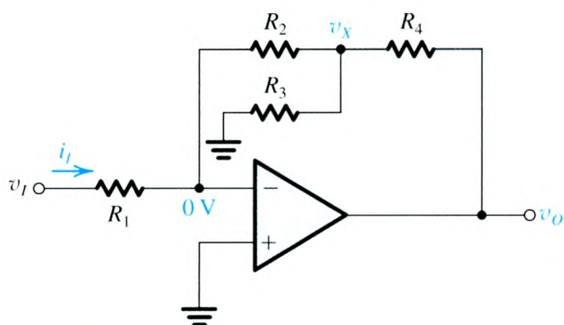


Figure P2.32

***2.33** The circuit in Fig. P2.33 can be considered to be an extension of the circuit in Fig. 2.8.

(a) Find the resistances looking into node 1, R_1; node 2, R_2; node 3, R_3; and node 4, R_4.
(b) Find the currents I_1, I_2, I_3, and I_4, in terms of the input current I.
(c) Find the voltages at nodes 1, 2, 3, and 4, that is, V_1, V_2, V_3, and V_4 in terms of (IR).

2.34 The circuit in Fig. P2.34 utilizes an ideal op amp.

(a) Find I_1, I_2, I_3, I_L, and V_x.
(b) If V_O is not to be lower than -8 V, find the maximum allowed value for R_L.

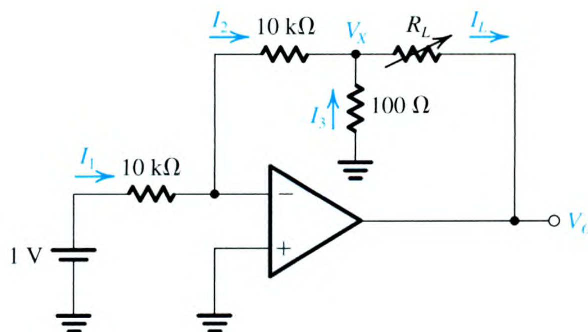


Figure P2.34

(c) If R_L is varied in the range 100 Ω to 500 kΩ, what is the corresponding change in I_L and in V_O?

D 2.35 Use the circuit in Fig. P2.34 as an inspiration to design a circuit that supplies a constant current I_L of 3.1 mA to a variable resistance R_L. Assume the availability of a 1.5-V battery and design so that the current drawn from the battery is 0.1 mA. For the smallest resistance in the circuit, use 500 Ω. If the op amp saturates at ±10 V, what is the maximum value that R_L can have while the current source supplying it operates properly?

VE 2.2 **D 2.36** Assuming the op amp to be ideal, it is required to design the circuit shown in Fig. VE 2.2 to implement a current amplifier with gain i_L/i_I = 11 A/A.

(a) Find the required value for R.
(b) What are the input and the output resistance of this current amplifier?
(c) If R_L = 1 kΩ and the op amp operates in an ideal manner as long as v_O is in the range ±12 V, what range of i_I is possible?

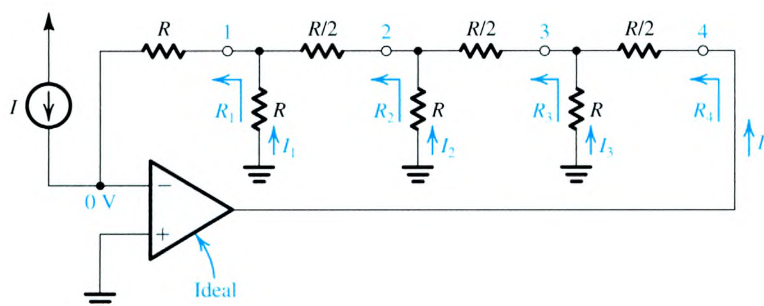


Figure P2.33

(d) If the amplifier is fed with a current source having a current of 0.2 mA and a source resistance of 10 kΩ, find i_L.

D 2.37 Design the circuit shown in Fig. P2.37 to have an input resistance of 100 kΩ and a gain that can be varied from −1 V/V to −100 V/V using the 100-kΩ potentiometer R_4. What voltage gain results when the potentiometer is set exactly at its middle value?

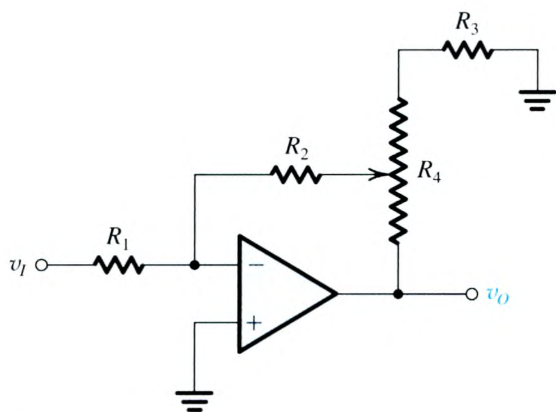


Figure P2.37

2.38 A weighted summer circuit using an ideal op amp as shown in Fig. 2.10 has three inputs using 20-kΩ resistors and a feedback resistor of 50 kΩ. A signal v_1 is connected to two of the inputs while a signal v_2 is connected to the third. Express v_O in terms of v_1 and v_2. If $v_1 = 1$ V and $v_2 = -1$ V, what is v_O?

2.39 Design an op-amp circuit to provide an output $v_O = -[2v_1 + (v_2/2)]$. Choose the lowest possible values of resistors for which the input current (from each input signal source) does not exceed 50 μA for 1-V input signals.

D 2.40 Use the scheme illustrated in Fig. 2.10 to design an op-amp circuit with inputs v_1, v_2, and v_3, whose output is $v_O = -(1v_1 + 4v_2 + 6v_3)$ using the smallest possible resistors but no smaller than 1 kΩ.

D 2.41 An ideal op amp is connected in the weighted summer configuration of Fig. 2.10. The feedback resistor $R_f = 100$ kΩ, and eight 100-kΩ resistors are connected to the inverting input terminal of the op amp. Show, by sketching the various circuit configurations, how this basic circuit can be used to implement the following functions:

(a) $v_O = -(v_1 + 2v_2 + 5v_3)$
(b) $v_O = -(v_1 + v_2 + 2v_3 + 4v_4)$
(c) $v_O = -(2v_1 + 6v_2)$
(d) $v_O = -8v_1$

In each case find the input resistance seen by each of the signal sources supplying v_1, v_2, v_3, and v_4. Suggest at least two additional summing functions that you can realize with this circuit. How would you realize a summing coefficient that is 0.5?

D 2.42 Give a circuit, complete with component values, for a weighted summer that shifts the dc level of a sine-wave signal of 3 sin(ωt) V from zero to −3 V. Assume that in addition to the sine-wave signal you have a dc reference voltage of 1.5 V available. Sketch the output signal waveform.

D 2.43 Use two ideal op amps and resistors to implement the summing function

$$v_O = v_1 + 3v_2 - 2v_3 - 5v_4$$

***2.44** Figure P2.44 shows a circuit for a digital-to-analog converter (DAC). The circuit accepts a 4-bit input binary word $a_3a_2a_1a_0$, where a_0, a_1, a_2, and a_3 take the values of 0 or 1, and it provides an analog output voltage v_O proportional to the value of the digital input. Each of the bits of the input word controls the correspondingly numbered switch. For instance, if a_2 is 0 then switch S_2 connects the 20-kΩ resistor to ground, while if a_2 is 1 then S_2 connects the 20-kΩ resistor to the +5-V power supply. Show that v_O is given by

$$v_O = -\frac{R_f}{16}[2^0a_0 + 2^1a_1 + 2^2a_2 + 2^3a_3]$$

where R_f is in kilohms. Find the value of R_f so that v_O ranges from 0 to −5 volts.

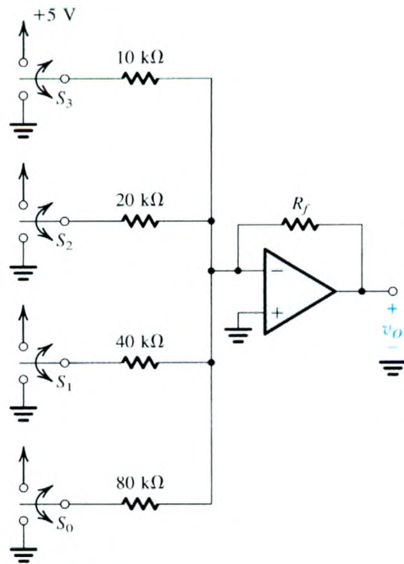


Figure P2.44

Section 2.3: The Noninverting Configuration

D 2.45 Given an ideal op amp to implement designs for the following closed-loop gains, what values of resistors (R_1, R_2) should be used? Use at least one 10-kΩ resistor as the smallest resistor in your design.

(a) +5 V/V
(b) +10 V/V
(c) +21 V/V
(d) +100 V/V

D 2.46 Design a circuit based on the topology of the noninverting amplifier to obtain a gain of +1.5 V/V, using only 10-kΩ resistors. Note that there are two possibilities. Which of these can be easily converted to have a gain of either +1.0 V/V or +2.0 V/V simply by short-circuiting a single resistor in each case?

D 2.47 Figure P2.47 shows a circuit for an analog voltmeter of very high input resistance that uses an inexpensive moving-coil meter. The voltmeter measures the voltage V applied between the op amp's positive-input terminal and ground. The moving-coil meter deflects a small mechanical arm in response to a current. Assuming it produces full-scale deflection when the current passing through it is 100 μA, find the value of R such that a full-scale reading is obtained when V is +10 V. Does the value of the meter resistance affect the voltmeter calibration?

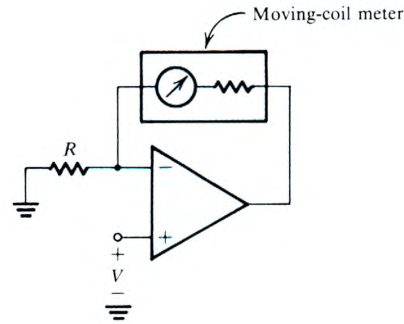


Figure P2.47

D*2.48 (a) Use superposition to show that the output of the circuit in Fig. P2.48 is given by

$$v_O = -\left[\frac{R_f}{R_{N1}}v_{N1} + \frac{R_f}{R_{N2}}v_{N2} + \cdots + \frac{R_f}{R_{Nn}}v_{Nn}\right] + \left[1 + \frac{R_f}{R_N}\right]\left[\frac{R_P}{R_{P1}}v_{P1} + \frac{R_P}{R_{P2}}v_{P2} + \cdots + \frac{R_P}{R_{Pn}}v_{Pn}\right]$$

where $R_N = R_{N1} \| R_{N2} \| \cdots \| R_{Nn}$, and

$$R_P = R_{P1} \| R_{P2} \| \cdots \| R_{Pn} \| R_{P0}$$

(b) Design a circuit to obtain

$$v_O = -4v_{N1} + v_{P1} + 3v_{P2}$$

The smallest resistor used should be 10 kΩ.

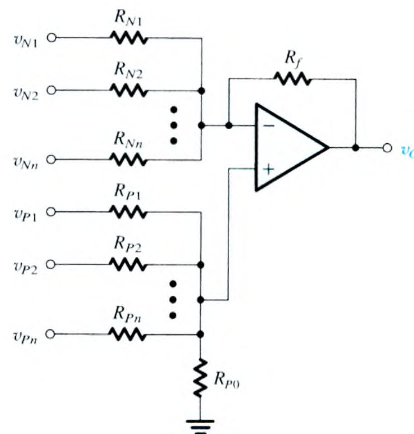


Figure P2.48

D*2.49 Design a circuit, using one ideal op amp, whose output is $v_O = v_{I1} + 2v_{I2} - 9v_{I3} + 4v_{I4}$. (*Hint:* Use a structure similar to that shown in general form in Fig. P2.48.)

2.50 Derive an expression for the voltage gain, v_O/v_I, of the circuit in Fig. P2.50.

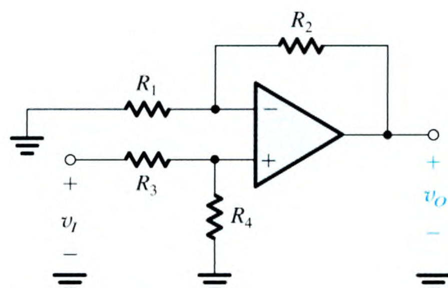


Figure P2.50

2.51 For the circuit in Fig. P2.51, use superposition to find v_O in terms of the input voltages v_1 and v_2. Assume an ideal op amp. For

$$v_1 = 10\sin(2\pi \times 60t) - 0.1\sin(2\pi \times 5000t), \text{ volts}$$
$$v_2 = 10\sin(2\pi \times 60t) + 0.1\sin(2\pi \times 5000t), \text{ volts}$$

find v_O.

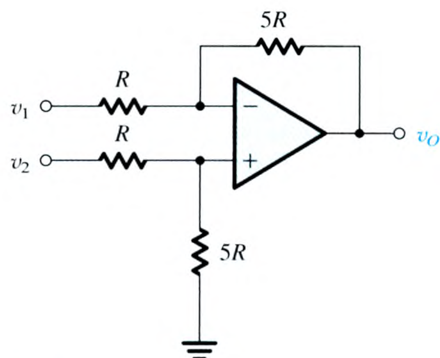


Figure P2.51

D 2.52 The circuit shown in Fig. P2.52 utilizes a 50-kΩ potentiometer to realize an adjustable-gain amplifier. Derive an expression for the gain as a function of the potentiometer setting x. Assume the op amp to be ideal. What is the range of gains obtained? Show how to add a fixed resistor so that the gain range can be 1 to 21 V/V. What should the resistor value be?

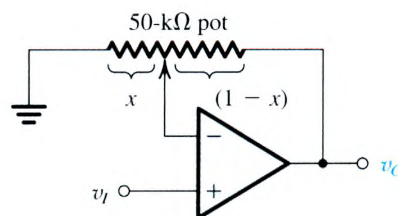


Figure P2.52

D 2.53 Given the availability of resistors of value 1 kΩ and 10 kΩ only, design a circuit based on the noninverting configuration to realize a gain of +10 V/V. What is the input resistance of your amplifier?

2.54 It is required to connect a 10-V source with a source resistance of 1 MΩ to a 1-kΩ load. Find the voltage that will appear across the load if:

(a) The source is connected directly to the load.
(b) An inverting amplifier utilizing 100-kΩ resistors and an ideal op amp to realize a gain of −1 V/V is inserted between the source and load.
(c) A unity-gain op-amp buffer is inserted between the source and load.

In each case find the load current and the current supplied by the source. Where does the load current come from in cases (b) and (c)?

2.55 Derive an expression for the gain of the voltage follower of Fig. 2.14, assuming the op amp to be ideal except for having a finite gain A. Calculate the value of the closed-loop gain for A = 60 dB, 40 dB, and 20 dB. In each case find the percentage error in gain magnitude from the nominal value of unity.

2.56 A 1-V peak sinusoid is applied to a voltage follower. Assuming the op amp to be ideal, what is the differential signal appearing at the op-amp input terminals? What is the common-mode signal appearing at the op-amp inputs? How does this compare with the differential and common-mode op-amp input signals arising in an inverting configuration designed for a gain of −1?

2.57 Complete the following table for feedback amplifiers created using one ideal op amp. Note that R_{in} signifies input resistance and R_1 and R_2 are feedback-network resistors as labeled in the inverting and noninverting configurations.

Case	Gain	R_{in}	R_1	R_2
a	−10 V/V	20 kΩ		
b	−4 V/V		100 kΩ	
c	−2 V/V			20 kΩ
d	+1 V/V	∞		
e	+25 V/V		100 kΩ	
f	+9 V/V			40 kΩ
g	−0.5 V/V	10 kΩ		

D 2.58 You wish to select an op amp for use in a noninverting configuration having a nominal gain of +20 V/V. Using Eq. (2.12), find the minimum op-amp gain A, expressed in decibels, required to ensure the actual closed-loop gain is within 1% of its nominal value.

D 2.59 A noninverting op-amp circuit with nominal gain of 10 V/V uses an op amp with open-loop gain of 100 V/V and a lowest-value resistor of 10 kΩ. What closed-loop gain actually results? With what value resistor can which resistor be shunted to achieve the nominal gain? If in the manufacturing process, an op amp of gain 200 V/V were used, what closed-loop gain would result in each case (the uncompensated one, and the compensated one)?

2.60 Use Eq. (2.11) to show that if the reduction in the closed-loop gain G from the nominal value $G_0 = 1 + R_2/R_1$ is to be kept less than x% of G_0, then the open-loop gain of the op amp must exceed G_0 by at least a factor $F = (100/x) - 1 \simeq 100/x$. Find the required F for $x = 0.01, 0.1, 1,$ and 10. Utilize these results to find for each value of x the minimum required open-loop gain to obtain closed-loop gains of 1, 10, 10^2, 10^3, and 10^4 V/V.

2.61 For each of the following combinations of op-amp open-loop gain A and nominal closed-loop gain G_0, calculate the actual closed-loop gain G that is achieved. Also, calculate the percentage by which $|G|$ falls short of the nominal gain magnitude $|G_0|$.

Case	G_0 (V/V)	A (V/V)
a	−1	10
b	+1	10
c	−5	100
d	+2	10
e	−10	100
f	−10	1000
g	+1	5

2.62 Figure P2.62 shows a circuit that provides an output voltage v_O whose value can be varied by turning the wiper of the 100-kΩ potentiometer. Find the range over which v_O can be varied. If the potentiometer is a "20-turn" device, find the change in v_O corresponding to each turn of the pot.

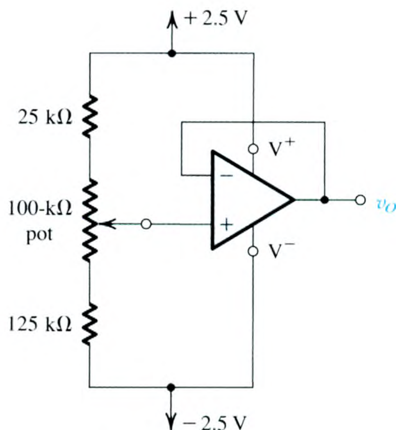


Figure P2.62

Section 2.4: Difference Amplifiers

2.63 Find the voltage gain v_O/v_{Id} for the difference amplifier of Fig. 2.16 for the case $R_1 = R_3 = 5$ kΩ and $R_2 = R_4 = 50$ kΩ. What is the differential input resistance R_{id}? If the two key resistance ratios (R_2/R_1) and (R_4/R_3) are different from each other by 1%, what do you expect the common-mode gain A_{cm} to be? Also, find the CMRR in this case. Neglect the effect of the ratio mismatch on the value of A_d.

D 2.64 Using the difference amplifier configuration of Fig. 2.16 and assuming an ideal op amp, design the circuit to provide the following differential gains. In each case, the differential input resistance should be 10 kΩ.

(a) 1 V/V
(b) 5 V/V
(c) 10 V/V
(d) 25 V/V

2.65 For the circuit shown in Fig. P2.65, express v_O as a function of v_1 and v_2. What is the input resistance seen by v_1 alone? By v_2 alone? By a source connected between the two input terminals? By a source connected to both input terminals simultaneously?

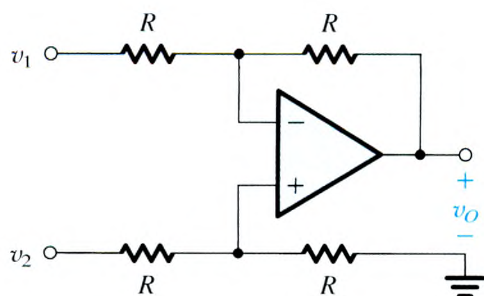


Figure P2.65

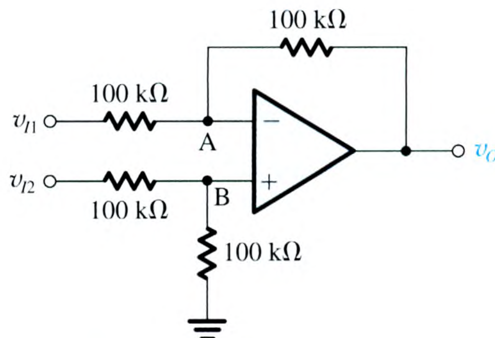


Figure P2.68

2.66 Consider the difference amplifier of Fig. 2.16 with the two input terminals connected together to an input common-mode signal source. For $R_2/R_1 = R_4/R_3$, show that the input common-mode resistance is $(R_3 + R_4) \| (R_1 + R_2)$.

***2.67** For the difference amplifier of Fig. 2.16, show that if each resistor has a tolerance of $\pm 100\epsilon\%$ (i.e., for, say, a 5% resistor, $\epsilon = 0.05$) then the worst-case CMRR is given approximately by

$$\text{CMRR} \simeq 20\log\left[\frac{K+1}{4\epsilon}\right]$$

where K is the nominal (ideal) value of the ratios (R_2/R_1) and (R_4/R_3). Calculate the value of worst-case CMRR for an amplifier designed to have a differential gain of ideally 100 V/V, assuming that the op amp is ideal and that 1% resistors are used. What resistor tolerance is needed if a CMRR of 80 dB is required?

***2.68** (a) Find A_d and A_{cm} for the difference amplifier circuit shown in Fig. P2.68.

(b) If the op amp is specified to operate properly as long as the common-mode voltage at its positive and negative inputs falls in the range ±2.5 V, what is the corresponding limitation on the range of the input common-mode signal v_{Icm}? (This is known as the **common-mode range** of the differential amplifier.)

(c) The circuit is modified by connecting a 10-kΩ resistor between node A and ground, and another 10-kΩ resistor between node B and ground. What will now be the values of A_d, A_{cm}, and the input common-mode range?

D*2.69 To obtain a high-gain, high-input-resistance difference amplifier, the circuit in Fig. P2.69 employs positive feedback, in addition to the negative feedback provided by the resistor R connected from the output to the negative input of the op amp. Specifically, a voltage divider (R_5, R_6) connected across the output feeds a fraction β of the output, that is, a voltage βv_O, back to the positive-input terminal of the op amp through a resistor R. Assume that R_5 and R_6 are much smaller than R so that the current through R is much lower than the current in the voltage divider, with the result that $\beta \simeq R_6/(R_5 + R_6)$. Show that the differential gain is given by

$$A_d \equiv \frac{v_O}{v_{Id}} = \frac{1}{1-\beta}$$

(*Hint:* Use superposition.)

Design the circuit to obtain a differential gain of 10 V/V and differential input resistance of 2 MΩ. Select values for R, R_5, and R_6, such that $(R_5 + R_6) \leq R/100$.

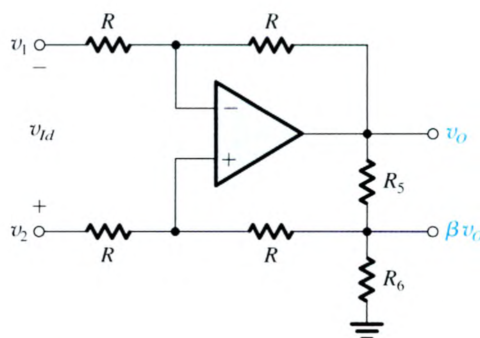


Figure P2.69

***2.70** Figure P2.70 shows a modified version of the difference amplifier. The modified circuit includes a resistor R_G,

which can be used to vary the gain. Show that the differential voltage gain is given by

$$\frac{v_O}{v_{Id}} = -2\frac{R_2}{R_1}\left[1 + \frac{R_2}{R_G}\right]$$

(*Hint:* The virtual short circuit at the op-amp input causes the current through the R_1 resistors to be $v_{Id}/2R_1$).

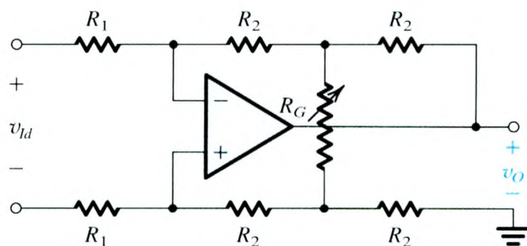


Figure P2.70

***2.71** For the differential amplifier of Fig. P2.71, derive an expression for the output, v_O, in terms of the differential input, $v_{Id} = v_{I2} - v_{I1}$, and the potentiometer setting x. Note that the circuit acts as a difference amplifier whose output can be adjusted by the potentiometer setting. What is the change in output voltage when x is changed from 0 to 1? How may this range be increased?

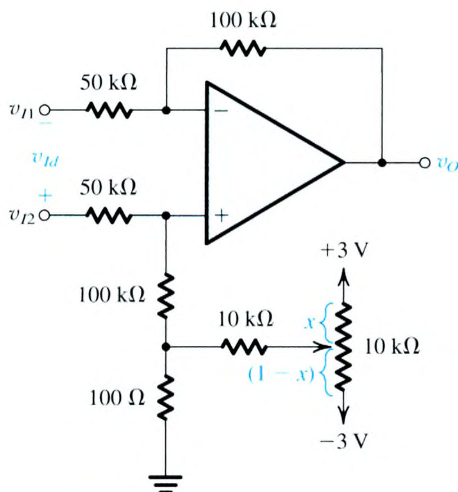


Figure P2.71

D*2.72 The circuit shown in Fig. P2.72 is a representation of a versatile IC known as a differential amplifier module available from different manufacturers under various part numbers such as INA105 and AMP03. It consists of an op amp and precision resistors. The circuit can be configured for a variety of applications by the appropriate connection of terminals A, B, C, D, and O.

(a) Show how the circuit can be used to implement a difference amplifier of unity gain.
(b) Show how the circuit can be used to implement single-ended amplifiers with gains:
  (i) −1 V/V
  (ii) +1 V/V
  (iii) +2 V/V
  (iv) +1/2 V/V

Avoid leaving a terminal open-circuited, for such a terminal may act as an "antenna," picking up interference and noise through capacitive coupling. Rather, find a convenient node to connect such a terminal in a redundant way. When more than one circuit implementation is possible, comment on the relative merits of each, taking into account such considerations as dependence on component matching and input resistance.

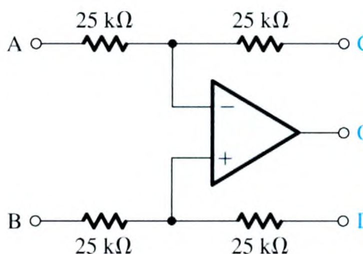


Figure P2.72

2.73 Consider the instrumentation amplifier of Fig. 2.20(b) with a common-mode input voltage of +1 V (dc) and a differential input signal of 100-mV peak-to-peak sine wave. Let $2R_1 = 2\ \text{k}\Omega$, $R_2 = 50\ \text{k}\Omega$, $R_3 = R_4 = 10\ \text{k}\Omega$. Find the voltage at every node in the circuit.

2.74 In the instrumentation amplifier of Fig. 2.20(a), one of the resistors R_2 is replaced by a slightly different value $R_2 + \Delta R$. Find an expression for the resulting gain from a common-mode input at v_{I1} and v_{I2} to the output v_O. Make this gain zero by modifying the value of one of the resistors R_1.

2.75 (a) Consider the instrumentation amplifier circuit of Fig. 2.20(a). If the op amps are ideal except that their outputs saturate at ±5 V, in the manner shown in Fig. 1.14, find the maximum allowed input common-mode signal for the case $R_1 = 1\ \text{k}\Omega$ and $R_2 = 100\ \text{k}\Omega$.
(b) Repeat (a) for the circuit in Fig. 2.20(b), and comment on the difference between the two circuits.

2.76 (a) Expressing v_{I1} and v_{I2} in terms of differential and common-mode components, find v_{O1} and v_{O2} in the circuit in Fig. 2.20(a) and hence find their differential component $v_{O2} - v_{O1}$ and their common-mode component $\frac{1}{2}(v_{O1} + v_{O2})$. Now find the differential gain and the common-mode gain of the first stage of this instrumentation amplifier and hence the CMRR.
(b) Repeat for the circuit in Fig. 2.20(b), and comment on the difference between the two circuits.

***2.77** For an instrumentation amplifier of the type shown in Fig. 2.20(b), a designer proposes to make $R_2 = R_3 = R_4 = 100\ \text{k}\Omega$, and $2R_1 = 10\ \text{k}\Omega$. For ideal components, what difference-mode gain, common-mode gain, and CMRR result? Reevaluate the worst-case values for these for the situation in which all resistors are specified as ±1% units. Repeat the latter analysis for the case in which $2R_1$ is reduced to 1 kΩ. What do you conclude about the effect of the gain of the first stage on CMRR? (*Hint:* Eq. (2.19) can be used to evaluate A_{cm} of the second stage.)

D 2.78 Design the instrumentation-amplifier circuit of Fig. 2.20(b) to realize a differential gain, variable in the range 2 to 50, utilizing a 100-kΩ pot as variable resistor.

SIM 2.79 *The circuit shown in Fig. P2.79 is intended to supply a voltage to floating loads (those for which both terminals are ungrounded) while making greatest possible use of the available power supply.

(a) Assuming ideal op amps, sketch the voltage waveforms at nodes B and C for a 1-V peak-to-peak sine wave applied at A. Also sketch v_O.
(b) What is the voltage gain v_O/v_I?

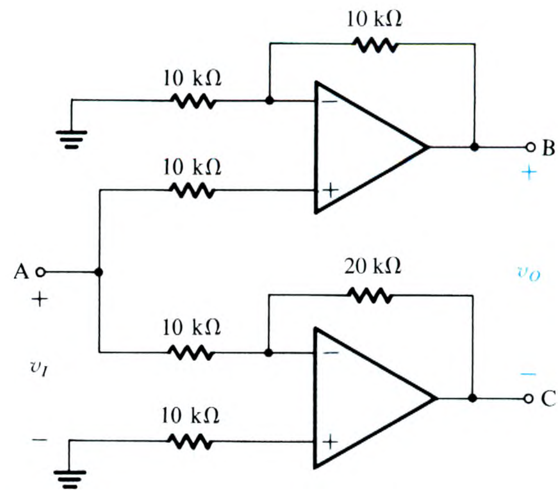


Figure P2.79

(c) Assuming that the op amps operate from ±5-V power supplies and that their output saturates at ±4 V (in the manner shown in Fig. 1.14), what is the largest sine-wave input that can be accommodated? Specify both its peak-to-peak and rms values.

***2.80** The two circuits in Fig. P2.80 are intended to function as voltage-to-current converters; that is, they supply the load impedance Z_L with a current proportional to v_I and independent of the value of Z_L. Show that this is indeed the case, and find for each circuit i_O as a function of v_I. Comment on the differences between the two circuits.

Section 2.5: Integrators and Differentiators

2.81 A Miller integrator incorporates an ideal op amp, a resistor R of 10 kΩ, and a capacitor C of 1 nF. A sine-wave signal is applied to its input.

(a) At what frequency (in Hz) are the input and output signals equal in amplitude?
(b) At that frequency, how does the phase of the output sine wave relate to that of the input?
(c) If the frequency is lowered by a factor of 10 from that found in (a), by what factor does the output voltage change, and in what direction (smaller or larger)?

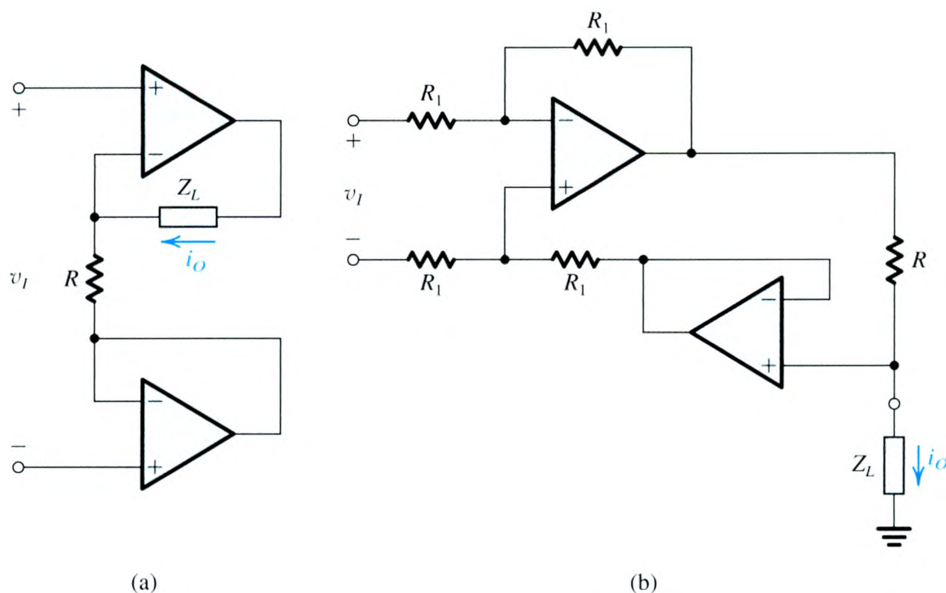


Figure P2.80

(d) What is the phase relation between the input and output in situation (c)?

D 2.82 Design a Miller integrator with a time constant of 1 ms and an input resistance of 100 kΩ. A dc voltage of −1 volt is applied at the input at time 0, at which moment $v_O = -5$ V. How long does it take the output to reach 0 V? +5 V?

2.83 An op-amp-based inverting integrator is measured at 10 kHz to have a voltage gain of -10^4 V/V. At what frequency is its gain reduced to −1 V/V? What is the integrator time constant?

2.84 An inverting integrator accepts an input pulse of amplitude +100 mV, 1 ms in duration. It draws 10 μA from the input voltage source, and its output changes by −1 V. Find the values of R and C and sketch the integrator magnitude response, labeling the integrator frequency.

D 2.85 Design a Miller integrator whose input resistance is 1 kΩ and unity-gain frequency is 10 MHz. What components are needed? For long-term stability, a feedback resistor is introduced across the capacitor to limit the dc gain to 40 dB. What is its value? What is the associated lower 3-dB frequency? Sketch and label the output that results with a 0.1-μs, 1-V positive-input pulse (initially at 0 V) with (a) no dc stabilization (but with the output initially at 0 V) and (b) the feedback resistor connected.

***2.86** A Miller integrator whose input and output voltages are initially zero and whose time constant is 0.1 ms is driven by the signal shown in Fig. P2.86. Sketch and label the output waveform that results. Indicate what happens if the input levels are ±1 V, with the time constant the same (0.1 ms) and with the time constant raised to 1 ms.

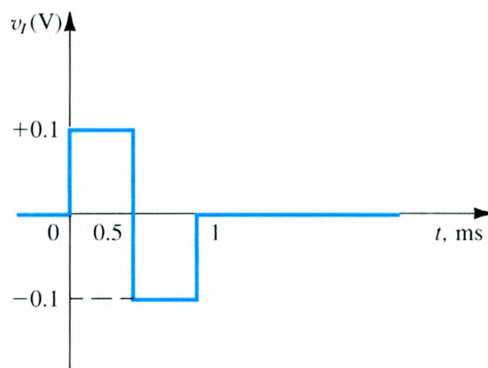


Figure P2.86

2.87 Consider a Miller integrator having a time constant of 0.2 ms and an output that is initially zero, when fed with a string of pulses of 10-μs duration and 1-V amplitude rising from 0 V (see Fig. P2.87). Sketch and label the output waveform resulting. How many pulses are required for an output voltage of 1 V?

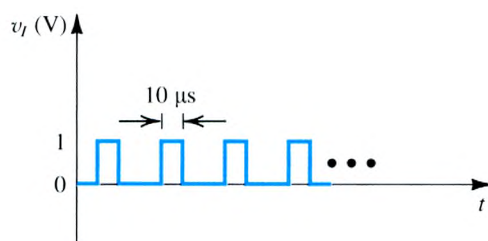


Figure P2.87

D 2.88 Figure 2.23 shows a circuit that performs a low-pass STC function and is known as a first-order low-pass active filter. Design the circuit to obtain an input resistance of 10 kΩ, a dc gain of 20 dB, and a 3-dB frequency of 200 kHz. Sketch the circuit's magnitude response indicating the frequency where it is 0 dB.

***2.89** Show that a Miller integrator implemented with an op amp with a finite open-loop gain A_0 has a low-pass STC transfer function. What is the pole frequency of the STC function? How does this compare with the pole frequency of the ideal integrator? If an ideal Miller integrator is fed with a −1-V pulse signal with a width $T = CR$, what will the output voltage be at $t = T$? Assume that at $t = 0, v_O = 0$. Repeat for an integrator with an op amp having $A_0 = 1000$.

2.90 An op-amp differentiator with 1-ms time constant is driven by the rate-controlled step shown in Fig. P2.90. Assuming v_O to be zero initially, sketch and label its waveform.

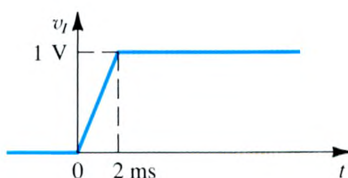


Figure P2.90

2.91 Use an ideal op amp to design a differentiator circuit having a time constant of 10^{-5} s using a 1-nF capacitor. Sketch the magnitude and phase response of the circuit, indicating the frequency at which the magnitude response equals 0 dB. A series input resistor is added to limit the gain magnitude at high frequencies to 100 V/V. Sketch the magnitude and phase responses of this modified circuit on the same axes as the ideal differentiator. How does the phase response change at high frequencies?

D 2.92 Figure P2.92 shows a first-order high-pass active filter that has a single time-constant response. Derive the transfer function and show that the high-frequency gain is $(-R_2/R_1)$ and the 3-dB frequency $\omega_0 = 1/CR_1$. Design the circuit to obtain a high-frequency input resistance of 2 kΩ, a high-frequency gain of 40 dB, and a 3-dB frequency of 2 MHz. At what frequency does the magnitude of the transfer function equal unity?

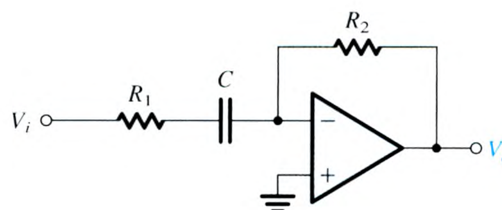


Figure P2.92

D2.93** Derive the transfer function of the circuit in Fig. P2.93 (for an ideal op amp) and show that it can be written in the form

$$\frac{V_o}{V_i} = \frac{-R_2/R_1}{[1 + (\omega_1/j\omega)][1 + j(\omega/\omega_2)]}$$

where $\omega_1 = 1/C_1R_1$ and $\omega_2 = 1/C_2R_2$. Assuming that the circuit is designed such that $\omega_2 \gg \omega_1$, find approximate expressions for the transfer function in the following frequency regions:

(a) $\omega \ll \omega_1$
(b) $\omega_1 \ll \omega \ll \omega_2$
(c) $\omega \gg \omega_2$

Use these approximations to sketch a Bode plot for the magnitude response. Observe that the circuit performs as an amplifier whose gain rolls off at the low-frequency end in the manner of a high-pass STC network, and at the high-frequency end in the manner of a low-pass STC network. Design the circuit to provide a gain of 40 dB in

the "middle-frequency range," a low-frequency 3-dB point at 200 Hz, a high-frequency 3-dB point at 200 kHz, and an input resistance (at $\omega \gg \omega_1$) of 2 kΩ.

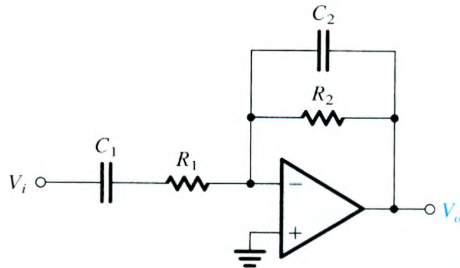


Figure P2.93

Section 2.6: DC Imperfections

2.94 An op amp wired in the inverting configuration with the input grounded, having $R_2 = 20\ \text{k}\Omega$ and $R_1 = 2\ \text{k}\Omega$, has an output dc voltage of −80 mV. If the input bias current is known to be very small, find the input offset voltage.

2.95 A noninverting amplifier with a gain of 25 uses an op amp having an input offset voltage of ±2 mV. Find the output when the input is 0.1 sin ωt volts.

2.96 A noninverting amplifier with a closed-loop gain of 100 is designed using an op amp having an input offset voltage of 3 mV and output saturation levels of ±3 V. What is the maximum amplitude of the sine wave that can be applied at the input without the output clipping? If the amplifier is capacitively coupled in the manner indicated in Fig. 2.37, what would the maximum possible amplitude be?

2.97 An op amp connected in a closed-loop inverting configuration having a gain of 46 dB and using relatively small-valued resistors is measured with input grounded to have a dc output voltage of −0.2 V. What is its input offset voltage? Prepare an offset-voltage-source sketch resembling that in Fig. 2.29. Be careful of polarities.

2.98 A particular inverting amplifier with nominal gain of −100 V/V uses an imperfect op amp in conjunction with 100-kΩ and 10-MΩ resistors. The output voltage is found to be +1 V when measured with the input open and +0.5 V with the input grounded.

(a) What is the bias current of this amplifier? In what direction does it flow?
(b) Estimate the value of the input offset voltage.
(c) A 10-MΩ resistor is connected between the positive-input terminal and ground. With the input left floating (disconnected), the output dc voltage is measured to be −0.1 V. Estimate the input offset current.

D*2.99 A noninverting amplifier with a gain of +10 V/V using 100 kΩ as the feedback resistor operates from a 5-kΩ source. For an amplifier offset voltage of 0 mV, but with a bias current of 2 μA and an offset current of 0.2 μA, what range of outputs would you expect? Indicate where you would add an additional resistor to compensate for the bias currents. What does the range of possible outputs then become? A designer wishes to use this amplifier with a 15-kΩ source. In order to compensate for the bias current in this case, what resistor would you use? And where?

D 2.100 The circuit of Fig. 2.37 is used to create an ac-coupled noninverting amplifier with a gain of 100 V/V using resistors no larger than 100 kΩ. What values of R_1, R_2, and R_3 should be used? For a break frequency due to C_1 at 10 kHz, and that due to C_2 at 1 kHz, what values of C_1 and C_2 are needed?

***2.101** Consider the difference amplifier circuit in Fig. 2.16 driven by an ideal voltage source. Let $R_1 = R_3 = 100\ \text{k}\Omega$ and $R_2 = R_4 = 1\ \text{M}\Omega$. If the op amp has $V_{OS} = 5$ mV, $I_B =$ 100 nA, and $I_{OS} = 20$ nA, find the worst-case (largest) dc offset voltage at the output.

***2.102** The circuit shown in Fig. P2.102 uses an op amp having a ±3-mV offset. What is its output offset voltage? What does the output offset become with the input ac coupled through a capacitor C? If, instead, a large capacitor is placed in series with the 10-kΩ resistor, what does the output offset become?

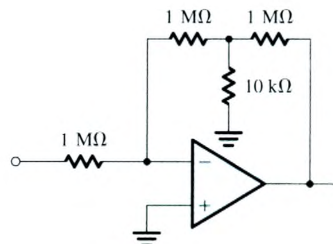


Figure P2.102

2.103 Using offset-nulling facilities provided for the op amp, a closed-loop amplifier with gain of +1000 is adjusted at 25°C to produce zero output with the input grounded. If the input offset-voltage drift is specified to be 20 μV/°C, what output would you expect at −40°C and at 85°C? While nothing can be said separately about the polarity of the output offset at either −40 or 85°C, what would you expect their relative polarities to be?


VE 2.3

2.104 An op amp is connected in a closed loop with gain of +100 utilizing a feedback resistor of 1 MΩ.

(a) If the input bias current is 200 nA, what output voltage results with the input grounded?
(b) If the input offset voltage is ±1 mV and the input bias current as in (a), what is the largest possible output that can be observed with the input grounded?
(c) If bias-current compensation is used, what is the value of the required resistor? If the offset current is no more than one-tenth the bias current, what is the resulting output offset voltage (due to offset current alone)?
(d) With bias-current compensation as in (c) in place, what is the largest dc voltage at the output due to the combined effect of offset voltage and offset current?

***2.105** An op amp intended for operation with a closed-loop gain of −100 V/V uses resistors of 10 kΩ and 1 MΩ with a bias-current-compensation resistor R_3. Assuming an ideal voltage source at the input, what should the value of R_3 be? With input grounded, for one particular op amp the output offset voltage is found to be +0.30 V. Estimate the input offset current assuming zero input offset voltage. In this case, if we know the input offset current is one-tenth the bias current what value of R_3 will result in zero output offset voltage?

2.106 A Miller integrator with $R = 10\,\text{k}\Omega$ and $C = 10\,\text{nF}$ is implemented by using an op amp with $V_{OS} = 2\,\text{mV}$, $I_B = 0.1\,\mu\text{A}$, and $I_{OS} = 20\,\text{nA}$. To provide a finite dc gain, a 1-MΩ resistor is connected across the capacitor.

(a) To compensate for the effect of I_B, a resistor is connected in series with the positive-input terminal of the op amp. What should its value be?
(b) With the resistor of (a) in place, find the worst-case dc output voltage of the integrator when the input is grounded.

Section 2.7: Effect of Finite Open-Loop Gain and Bandwidth on Circuit Performance

2.107 The data in the following table apply to internally compensated op amps. Fill in the blank entries.

A_0	f_b (Hz)	f_t (Hz)
10^5	10^3	
10^4		10^8
	5×10^2	10^8
	20	10^7
2×10^5	10	

2.108 A measurement of the open-loop gain of an internally compensated op amp at very low frequencies shows it to be 106 dB. At 200 kHz, it is 40 dB. Estimate values for A_0, f_b, and f_t.

2.109 Measurements of the open-loop gain of a compensated op amp intended for high-frequency operation indicate that the gain is 4×10^3 at 100 kHz and 8×10^2 at 1 MHz. Estimate its 3-dB frequency, its unity-gain frequency, and its dc gain.

2.110 Measurements made on the internally compensated amplifiers listed below provide the dc gain and the frequency at which the gain has dropped by 20 dB. For each, what are the 3 dB and unity-gain frequencies?

(a) 2×10^5 V/V and 5×10^2 Hz
(b) 20×10^3 V/V and 1 kHz
(c) 1800 V/V and 0.1 MHz
(d) 60 dB and 10 MHz
(e) 200 V/mV and 25 kHz

SIM **2.111** An inverting amplifier with nominal gain of −50 V/V employs an op amp having a dc gain of 10^4 and a unity-gain frequency of 10^8 Hz. What is the 3-dB frequency f_{3dB} of the closed-loop amplifier? What is its gain at $0.1\,f_{3dB}$ and at $10\,f_{3dB}$?

2.112 A particular op amp, characterized by a gain–bandwidth product of 20 MHz, is operated with a closed-loop gain of +25 V/V. What 3-dB bandwidth results? At what frequency does the closed-loop amplifier exhibit a −6° phase shift? A −84° phase shift?

2.113 Find the f_t required for internally compensated op amps to be used in the implementation of closed-loop amplifiers with the following nominal dc gains and 3-dB bandwidths:

(a) −50 V/V; 200 kHz
(b) +50 V/V; 200 kHz
(c) +2 V/V; 5 MHz
(d) −2 V/V; 5 MHz
(e) −1000 V/V; 100 kHz

(f) +1 V/V; 20 MHz
(g) −1 V/V; 20 MHz

2.114 A noninverting op-amp circuit with a gain of 34 dB is found to have a 3-dB frequency of 200 kHz. For a particular system application, a bandwidth of 1 MHz is required. What is the highest gain available under these conditions?


VE 2.4

SIM **2.115** Consider a noninverting amplifier with gain of 5 utilizing an internally compensated op amp with $f_t = 2$ MHz. What is the 3-dB frequency of the amplifier? At what frequency is the gain of the amplifier 1% below its low-frequency magnitude? If the input is a 10-mV step, find the 10% to 90% rise time of the output voltage. (*Note:* The step response of STC low-pass networks is discussed in Appendix E. Specifically, note that the 10% to 90% rise time of a low-pass STC circuit with a time constant τ is 2.2τ.)

D*2.116 It is required to design a noninverting amplifier with a dc gain of 5. When a step voltage of 10 mV is applied at the input, it is required that the output be within 1% of its final value in at most 20 ns. What must the f_t of the op amp be? (*Note:* The step response of STC low-pass networks is discussed in Appendix E.)

D*2.117 This problem illustrates the use of cascaded closed-loop amplifiers to obtain an overall bandwidth greater than can be achieved using a single-stage amplifier with the same overall gain.

(a) Show that cascading two identical amplifier stages, each having a low-pass STC frequency response with a 3-dB frequency f_1, results in an overall amplifier with a 3-dB frequency given by
$$f_{3dB} = \sqrt{\sqrt{2}-1}f_1$$
(b) It is required to design a noninverting amplifier with a dc gain of 40 dB utilizing a single internally compensated op amp with $f_t = 2$ MHz. What is the 3-dB frequency obtained?
(c) Redesign the amplifier of (b) by cascading two identical noninverting amplifiers each with a dc gain of 20 dB. What is the 3-dB frequency of the overall amplifier? Compare this to the value obtained in (b) above.

D2.118** A designer, wanting to achieve a stable gain of 100 V/V with a 3-dB frequency above 5 MHz, considers her choice of amplifier topologies. What unity-gain frequency would a single operational amplifier require to satisfy her need? Unfortunately, the best available amplifier has an f_t of 50 MHz. How many such amplifiers connected in a cascade of identical noninverting stages would she need to achieve her goal? What is the 3-dB frequency of each stage? What is the overall 3-dB frequency?

2.119 Consider the use of an op amp with a unity-gain frequency f_t in the realization of:

(a) An inverting amplifier with dc gain of magnitude K.
(b) A noninverting amplifier with a dc gain of K.

In each case find the 3-dB frequency and the gain–bandwidth product (GBP ≡ |Gain| × f_{3dB}). Comment on the results.

****2.120** Consider an inverting summer with two inputs, V_1 and V_2, designed so that an ideal op amp provides $V_o = -(V_1 + 2V_2)$. If the op amp has a finite gain A, show that

$$V_o = -(V_1 + 2V_2)\frac{A/4}{1 + A/4}$$

Section 2.8: Large-Signal Operation of Op Amps

2.121 A particular op amp using ±15-V supplies operates linearly for outputs in the range −14 V to +14 V. If used in an inverting amplifier configuration of gain −100, what is the rms value of the largest possible sine wave that can be applied at the input without output clipping?

SIM **2.122** Consider an op amp connected in the inverting configuration to realize a closed-loop gain of −100 V/V utilizing resistors of 1 kΩ and 100 kΩ. A load resistance R_L is connected from the output to ground, and a low-frequency sine-wave signal of peak amplitude V_p is applied to the input. Let the op amp be ideal except that its output voltage saturates at ±10 V and its output current is limited to the range ±10 mA. This is the case for an ADA4077 op amp operating from ±11-V supplies.

(a) For $R_L = 2$ kΩ, what is the maximum possible value of V_p while an undistorted output sinusoid is obtained?
(b) Repeat (a) for $R_L = 200$ Ω.

(c) If it is desired to obtain an output sinusoid of 10-V peak amplitude, what minimum value of R_L is allowed?

SIM **2.123** An op amp having a slew rate of 1.2 V/μs (such as the ADA4077 operating from ±5 V supplies with a 2 kΩ load) is to be used in the unity-gain follower configuration, with input pulses that rise from 0 to 2 V. What is the shortest pulse that can be used while ensuring full-amplitude output? For such a pulse, describe the output resulting.

2.124 For operation with 10-V output pulses with the requirement that the sum of the rise and fall times represent only 20% of the pulse width (at half-amplitude), what is the slew-rate requirement for an op amp to handle pulses 20 μs wide? (*Note:* The rise and fall times of a pulse signal are usually measured between the 10%- and 90%-height points.)

***2.125** An inverting amplifier with a gain of −5 V/V and input resistance 20 kΩ uses an op amp with a slew rate of 1 V/μs, maximum output current of ±1 mA, and $f_t = 3$ MHz.

(a) At what frequency will the amplifier gain drop by 1 dB?
(b) At the frequency specified in (a), what is the maximum amplitude sinusoid for which the amplifier will avoid slew rate limiting?
(c) Under the conditions described in parts (a) and (b), what is the largest capacitive load that can be driven while staying within the maximum output current limits?

2.126 For an amplifier having a slew rate of 40 V/μs, what is the highest frequency at which a 2-V peak-to-peak sine wave can be produced at the output?

D*2.127 In designing with op amps one has to check the limitations on the voltage and frequency ranges of operation of the closed-loop amplifier, imposed by the op-amp finite bandwidth (f_t), slew rate (SR), and output saturation (V_{omax}). This problem illustrates the point by considering the use of an op amp with $f_t = 20$ MHz, SR = 10 V/μs, and $V_{omax} = 10$ V in the design of a noninverting amplifier with a nominal gain of 10. Assume a sine-wave input with peak amplitude V_i.

(a) If $V_i = 0.5$ V, what is the maximum frequency before the output distorts?
(b) If $f = 200$ kHz, what is the maximum value of V_i before the output distorts?
(c) If $V_i = 50$ mV, what is the useful frequency range of operation?
(d) If $f = 50$ kHz, what is the useful input voltage range?

CHAPTER 3

Semiconductors

IN THIS CHAPTER YOU WILL LEARN

- The basic properties of semiconductors and in particular silicon, which is the material used to make most of today's electronic circuits.
- How doping a pure silicon crystal dramatically changes its electrical conductivity, which is the fundamental idea underlying the use of semiconductors in the implementation of electronic devices.
- The two mechanisms by which current flows in semiconductors: drift and diffusion of charge carriers.
- The structure and operation of the *pn* junction; a basic semiconductor structure that implements the diode and plays a dominant role in transistors.

Introduction

Thus far we have dealt with electronic circuits, and notably amplifiers, as system building blocks. For instance, in Chapter 2 we learned how to use op amps to design interesting and useful circuits, taking advantage of the terminal characteristics of the op amp and without any knowledge of what is inside the op-amp package. Though interesting and motivating, this approach has limitations. Indeed, to become circuit designers, we have to go beyond this black-box or system-level abstraction and learn about the basic devices from which electronic circuits are assembled, namely, diodes (Chapter 4) and transistors (Chapters 5 and 6). These solid-state devices are made using semiconductor materials, predominantly silicon.

In this chapter, we introduce the properties and physics of semiconductors. The objective is to provide you with a basis for understanding the physical operation of diodes and transistors so that you may use them effectively in circuit design. Although many of the concepts presented in this chapter apply to semiconductor materials in general, our treatment is heavily biased toward silicon, the material used in the vast majority of microelectronic circuits. To complement the material presented here, Appendix A describes the integrated-circuit fabrication process. As explained in Appendix A, whether our circuit consists of a single transistor or is an **integrated circuit** containing more than 6 billion transistors, it is fabricated in a single silicon crystal, which gives rise to the name **monolithic circuit**. This chapter therefore begins with a study of the crystal structure of semiconductors and introduces the two types of charge carriers available for current conduction: electrons and holes. The most significant property of semiconductors is that their conductivity can be varied over a very wide range through the introduction of controlled amounts of impurity atoms into the semiconductor crystal in a process called **doping**. Doped semiconductors are discussed in Section 3.2. This is followed by the study in Section 3.3 of the two mechanisms for current flow in semiconductors, namely, carrier drift and carrier diffusion.

Armed with these basic semiconductor concepts, we spend the remainder of the chapter on the study of an important semiconductor structure: the *pn* junction. In addition to being essentially a diode, the *pn* junction is the basic element of the bipolar junction transistor (BJT, Chapter 6) and plays an important role in the operation of field-effect transistors (FETs, Chapter 5).

3.1 Intrinsic Semiconductors

As their name implies, semiconductors are materials whose conductivity lies between that of conductors, such as copper, and insulators, such as glass. There are two kinds of semiconductors: single-element semiconductors, such as germanium and silicon, which are in group IV in the periodic table; and compound semiconductors, such as gallium-arsenide, which are formed by combining elements from groups III and V or groups II and VI. Compound semiconductors are useful in special electronic circuit applications as well as in applications that involve light, such as light-emitting diodes (LEDs). One of the two elemental semiconductors, germanium, was used in the fabrication of very early transistors (late 1940s, early 1950s). It was quickly supplanted by silicon, on which today's

integrated-circuit technology is almost entirely based. For this reason, we will deal mostly with silicon devices throughout this book.[1]

A silicon atom has four valence electrons, and thus it requires another four to complete its outermost shell. This is achieved by sharing one of its valence electrons with each of its four neighboring atoms. Each pair of shared electrons forms a **covalent bond**. The result is that a crystal of pure or intrinsic silicon has a regular lattice structure, where the atoms are held in their position by the covalent bonds. Figure 3.1 shows a two-dimensional representation of such a structure.

At sufficiently low temperatures, approaching absolute zero (0 K), all the covalent bonds are intact and no electrons are available to conduct electric current. Thus, at such low temperatures, the intrinsic silicon crystal behaves as an insulator.

At room temperature, sufficient thermal energy exists to break some of the covalent bonds, a process known as thermal generation. As shown in Fig. 3.2, when a covalent bond is broken, an electron is freed. The **free electron** can wander away from its parent atom, and it becomes available to conduct electric current if an electric field is applied to the crystal. As the electron leaves its parent atom, it leaves behind a net positive charge, equal to the magnitude of the electron charge. Thus, an electrons from a neighboring atom may be attracted to this positive charge, and leaves its parent atom. This action fills up the "hole" that existed in the ionized atom but creates a new hole in the other atom. This process may repeat itself, with the result that we effectively have a positively charged carrier, or **hole**, moving through the silicon crystal structure and being available to conduct electric current. The charge of a hole is equal in magnitude to the charge of an electron. As temperature increases,

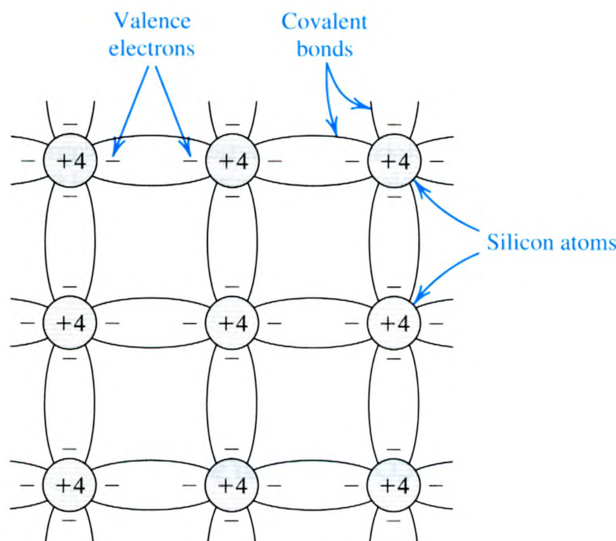


Figure 3.1 Two-dimensional representation of the silicon crystal. The circles represent the inner core of silicon atoms, with +4 indicating its positive charge of $+4q$, which is neutralized by the charge of the four valence electrons. Observe how the covalent bonds are formed by sharing of the valence electrons. At 0 K, all bonds are intact and no free electrons are available for current conduction.

[1] An exception is the gallium arsenide (GaAs) circuit, which though not covered here, is presented in some detail in the online student resources.

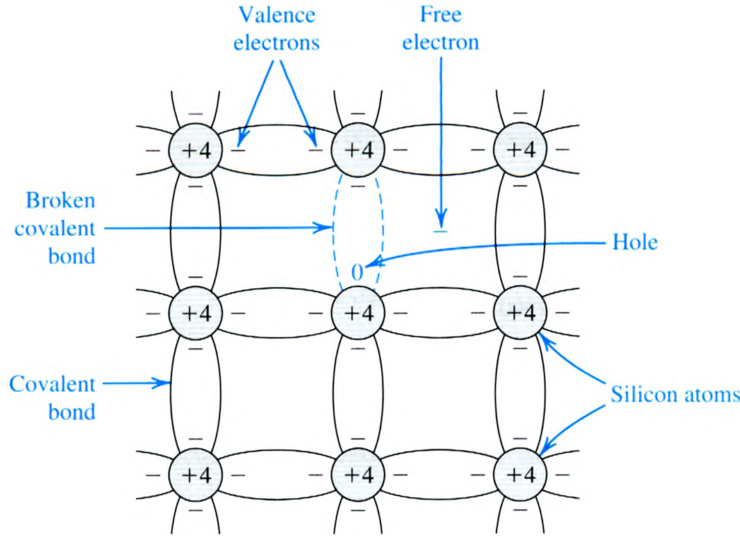


Figure 3.2 At room temperature, some of the covalent bonds are broken by thermal generation. Each broken bond gives rise to a free electron and a hole, both of which become available for current conduction.

more covalent bonds are broken and electron–hole pairs are generated. The increase in the numbers of free electrons and holes results in an increase in the conductivity of silicon.

Thermal generation results in free electrons and holes in equal numbers and hence equal concentrations, where concentration refers to the number of charge carriers per unit volume (cm^3). The free electrons and holes move randomly through the silicon crystal structure, and in the process some electrons may fill some of the holes. This process, called **recombination**, results in the disappearance of free electrons and holes. The recombination rate is proportional to the number of free electrons and holes, which in turn is determined by the thermal **generation** rate. The latter is a strong function of temperature. In thermal equilibrium, the recombination rate is equal to the generation rate, and we can conclude that the concentration of free electrons n is equal to the concentration of holes p,

$$n = p = n_i \tag{3.1}$$

where n_i denotes the number of free electrons and holes in a unit volume (cm^3) of intrinsic silicon at a given temperature. Results from semiconductor physics give n_i as

$$n_i = BT^{3/2}e^{-E_g/2kT} \tag{3.2}$$

where B is a material-dependent parameter that is $7.3 \times 10^{15} cm^{-3}K^{-3/2}$ for silicon; T is the temperature in K; E_g, a material parameter known as the **bandgap energy**, is 1.12 electron volt (eV) for silicon[2]; and k is Boltzmann's constant (8.62×10^{-5} eV/K). The bandgap energy E_g is the minimum energy required to break a covalent bond and thus generate an electron-hole pair.

[2]Note that 1 eV = 1.6×10^{-19} J.

Example 3.1

Calculate the value of n_i for silicon at room temperature ($T \simeq 300$ K).

Solution

Substituting the values given above in Eq. (3.2) provides

$$n_i = 7.3 \times 10^{15}(300)^{3/2}e^{-1.12/\left(2 \times 8.62 \times 10^{-5} \times 300\right)}$$

$$= 1.5 \times 10^{10}\text{carriers/cm}^3$$

Although this number seems large, to put it in context, note that silicon has 5×10^{22} atoms/cm^3. Thus at room temperature only one in about 5×10^{12} atoms is ionized and contributing a free electron and a hole!

Finally, it is useful for future purposes to express the product of the hole and free-electron concentration as

$$pn = n_i^2 \tag{3.3}$$

where for silicon at room temperature, $n_i \simeq 1.5 \times 10^{10}$/cm^3. As we will see shortly, this relationship extends to extrinsic or doped silicon as well.

LCDs, THE FACE OF ELECTRONICS

The existence of liquid crystals whose color could be changed by means of an external heat source was first reported in 1888 by an Austrian botanical physiologist. The LC idea lay dormant until the late 1940s, however. Subsequent developments in the field of solid-state electronics provided the technology to harness the technique in display media, with the first LCDs being demonstrated by RCA beginning in 1962. Today, LCDs are an essential component in many mobile devices as the interface to the world of electronics within. At the other end of the scale, large LCDs are used in flat-panel TVs, and very large LCDs are appearing as "dynamic" wallpaper in museum display settings.

EXERCISE

3.1 Calculate the intrinsic carrier density n_i for silicon at $T = 50$ K and 350 K.
Ans. 9.6×10^{-39}/cm^3; 4.15×10^{11}/cm^3

3.2 Doped Semiconductors

The intrinsic silicon crystal described above has equal concentrations of free electrons and holes, generated by thermal generation. These concentrations are far too small for silicon to conduct appreciable current at room temperature. Also, the carrier concentrations and hence the conductivity are strong functions of temperature, not a desirable property in an electronic device. Fortunately, a method was developed to change the carrier concentration in a semiconductor crystal substantially and in a precisely controlled manner. This process is known as doping, and the resulting silicon is referred to as **doped silicon**.

Doping involves introducing impurity atoms into the silicon crystal in sufficient numbers to substantially increase the concentration of either free electrons or holes but with little or no change in the crystal properties of silicon. To increase the concentration of free electrons, n, silicon is doped with an element with a valence of 5, such as phosphorus. The resulting doped silicon is then said to be of ***n* type**. To increase the concentration of holes, p, silicon is doped with an element having a valence of 3, such as boron, and the resulting doped silicon is said to be of ***p* type**.

Figure 3.3 shows a silicon crystal doped with phosphorus impurity. The dopant (phosphorus) atoms replace some of the silicon atoms in the crystal structure. Since the phosphorus atom has five electrons in its outer shell, four of these electrons form covalent bonds with the neighboring atoms, and the fifth electron becomes a free electron. Thus each phosphorus atom *donates* a free electron to the silicon crystal, and the phosphorus impurity is called a **donor**. It should be clear, though, that no holes are generated by this process. The net positive charge associated with the phosphorus atom is a **bound charge** that does not move through the crystal.

If the concentration of donor atoms is N_D, where N_D is usually much greater than n_i, the concentration of free electrons in the n-type silicon will be

$$n_n \simeq N_D \tag{3.4}$$

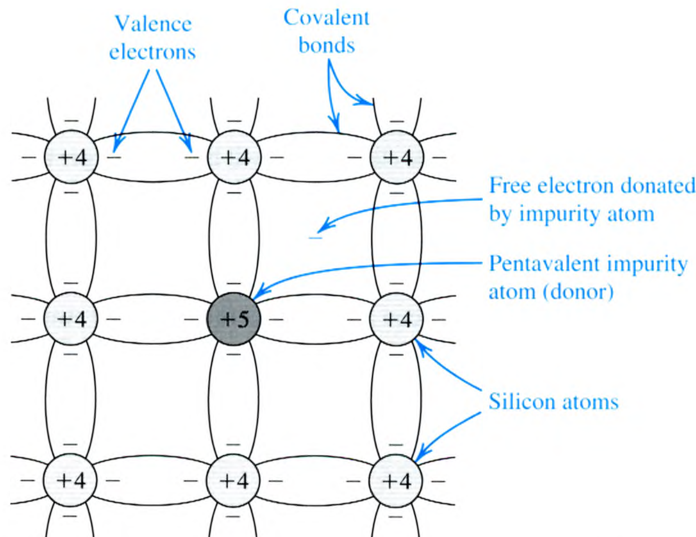


Figure 3.3 A silicon crystal doped by a pentavalent element. Each dopant atom donates a free electron and is thus called a donor. The doped semiconductor becomes n type.

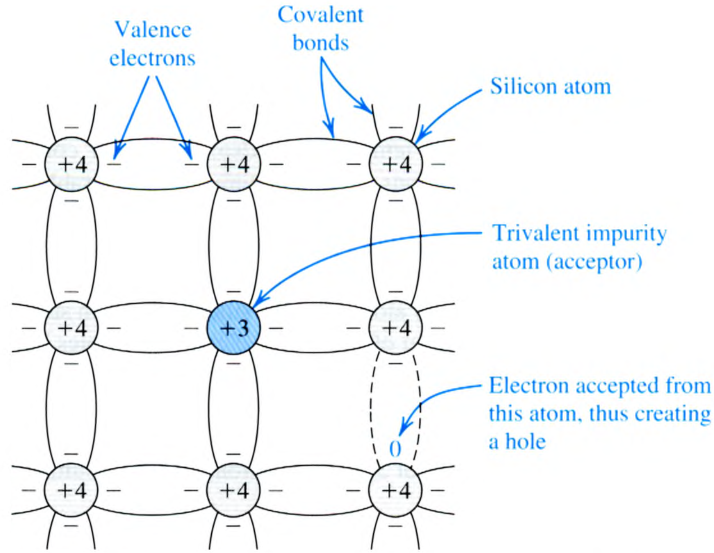


Figure 3.4 A silicon crystal doped with boron, a trivalent impurity. Each dopant atom gives rise to a hole, and the semiconductor becomes *p* type.

where the subscript n denotes n-type silicon. Thus n_n is determined by the doping concentration and not by temperature. This is not the case, however, for the hole concentration. All the holes in the n-type silicon are those generated by thermal ionization. Their concentration p_n can be found by noting that the relationship in Eq. (3.3) applies equally well for doped silicon, provided thermal equilibrium is achieved. Thus for n-type silicon

$$p_n n_n = n_i^2$$

Substituting for n_n from Eq. (3.4), we obtain for p_n

$$p_n \simeq \frac{n_i^2}{N_D} \tag{3.5}$$

Thus p_n will have the same dependence on temperature as that of n_i^2. Finally, we note that in n-type silicon the concentration of free electrons n_n will be much larger than that of holes. Hence electrons are said to be the **majority** charge carriers and holes the **minority** charge carriers in n-type silicon.

To obtain p-type silicon in which holes are the majority charge carriers, a trivalent impurity such as boron is used. Figure 3.4 shows a silicon crystal doped with boron. Note that the boron atoms replace some of the silicon atoms in the silicon crystal structure. Since each boron atom has three electrons in its outer shell, it **accepts** an electron from a neighboring atom, thus forming covalent bonds. The result is a hole in the neighboring atom and a bound negative charge at the **acceptor** (boron) atom. It follows that each acceptor atom provides a hole. If the acceptor doping concentration is N_A, where $N_A \gg n_i$, the hole concentration becomes

$$p_p \simeq N_A \tag{3.6}$$

where the subscript p denotes p-type silicon. Here the majority carriers are holes and their concentration is determined by N_A. The concentration of minority electrons can be found by

using the relationship

$$p_p n_p = n_i^2$$

and substituting for p_p from Eq. (3.6),

$$n_p \simeq \frac{n_i^2}{N_A} \tag{3.7}$$

Thus, the concentration of the minority electrons will have the same temperature dependence as that of n_i^2.

We wish to emphasize that a piece of n-type or p-type silicon is electrically neutral; the charge of the majority free carriers (electrons in the n-type and holes in the p-type silicon) is neutralized by the bound charges associated with the impurity atoms.

Example 3.2

Consider an n-type silicon for which the dopant concentration $N_D = 10^{17}/\text{cm}^3$. Find the electron and hole concentrations at $T = 300$ K.

Solution

The concentration of the majority electrons is

$$n_n \simeq N_D = 10^{17}/\text{cm}^3$$

The concentration of the minority holes is

$$p_n \simeq \frac{n_i^2}{N_D}$$

In Example 3.1 we found that at $T = 300$ K, $n_i = 1.5 \times 10^{10}/\text{cm}^3$. Thus,

$$p_n = \frac{\left(1.5 \times 10^{10}\right)^2}{10^{17}}$$
$$= 2.25 \times 10^3/\text{cm}^3$$

Observe that $n_n \gg n_i$ and that n_n is vastly higher than p_n.

EXERCISES

3.2 For the situation in Example 3.2, find the electron and hole concentrations at 350 K. You may use the value of n_i at $T = 350$ K found in Exercise 3.1.
Ans. $n_n = 10^{17}/\text{cm}^3$, $p_n = 1.72 \times 10^6/\text{cm}^3$

3.3 For a silicon crystal doped with boron, what must N_A be if at $T = 300$ K the electron concentration drops below the intrinsic level by a factor of 10^6?
Ans. $N_A = 1.5 \times 10^{16}/\text{cm}^3$

3.3 Current Flow in Semiconductors

There are two distinctly different mechanisms for the movement of charge carriers and hence for current flow in semiconductors: drift and diffusion.

3.3.1 Drift Current

When an electrical field E is established in a semiconductor crystal, holes are accelerated in the direction of E, and free electrons are accelerated in the direction opposite to that of E. This is illustrated in Fig. 3.5. The holes acquire a velocity

$$v_{p\text{-drift}} = \mu_p E \tag{3.8}$$

where μ_p is a constant called the **hole mobility**: It represents the ease by which holes move through the silicon crystal in response to the electrical field E. Since velocity has the units of centimeters per second and E has the units of volts per centimeter, we see from Eq. (3.8) that the mobility μ_p must have the units of centimeters squared per volt-second ($\text{cm}^2/\text{V}\cdot\text{s}$). For intrinsic silicon $\mu_p = 480\ \text{cm}^2/\text{V}\cdot\text{s}$.

The free electrons acquire a drift velocity

$$v_{n\text{-drift}} = -\mu_n E \tag{3.9}$$

where the result is negative because the electrons move in the direction opposite to E. Here μ_n is the **electron mobility**, which for intrinsic silicon is about $1350\ \text{cm}^2/\text{V}\cdot\text{s}$. Note that μ_n is about 2.5 times μ_p, signifying that electrons move with much greater ease than holes through the silicon crystal.

Let's now return to the single-crystal silicon bar shown in Fig. 3.5. Let the concentration of holes be p and that of free electrons n. We wish to calculate the current component due to the flow of holes. Consider a plane perpendicular to the x direction. The silicon bar has a cross-sectional area A in this direction. In one second, the hole charge that crosses that plane will be ($Aqpv_{p\text{-drift}}$) coulombs, where q is the magnitude of electron charge. This then must be the hole component of the drift current flowing through the bar,

$$I_{S,p} = Aqpv_{p\text{-drift}} \tag{3.10}$$

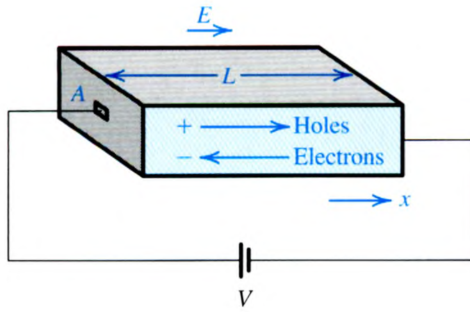


Figure 3.5 An electric field E established in a bar of silicon causes the holes to drift in the direction of E and the free electrons to drift in the opposite direction. Both the hole and electron drift currents are in the direction of E.

Substituting for $v_{p\text{-drift}}$ from Eq. (3.8), we obtain

$$I_{S,p} = Aqp\mu_p E$$

We are usually interested in the current density J_p, which is the current per unit cross-sectional area,

$$J_{S,p} = \frac{I_{S,p}}{A} = qp\mu_p E \tag{3.11}$$

The current due to the drift of free electrons can be found in a similar manner. By convention, the direction of current flow is in the direction of positive charge flow and opposite to the direction of negative charge flow. Thus, electrons drifting from right to left result in current flowing from left to right.

$$I_{S,n} = -Aqnv_{n\text{-drift}}$$

Substituting for $v_{n\text{-drift}}$ from Eq. (3.9), we obtain the current density $J_n = I_{S,n}/A$ as

$$J_{S,n} = qn\mu_n E \tag{3.12}$$

We can now find the total drift current density by summing $J_{S,p}$ and $J_{S,n}$ from Eqs. (3.11) and (3.12),

$$J_S = J_{S,p} + J_{S,n} = q(p\mu_p + n\mu_n)E \tag{3.13}$$

We can express this as

$$J = E/\rho \tag{3.14}$$

where the **resistivity** ρ is

$$\rho = \frac{1}{q(p\mu_p + n\mu_n)} \tag{3.15}$$

Eq. (3.14) is a form of Ohm's law and can be written alternately as

$$\rho = \frac{E}{J} \tag{3.16}$$

Thus the units of ρ are: $\dfrac{\text{V/cm}}{\text{A/cm}^2} = \Omega \cdot \text{cm}$.

We can find the resistance of the single-crystal silicon bar in Fig. 3.5 in terms of its resistivity, cross-sectional area, and length L:

$$R = \frac{\rho L}{A} \tag{3.17}$$

Example 3.3

Find the resistivity of (a) intrinsic silicon and (b) p-type silicon with $N_A = 10^{16}/\text{cm}^3$. Use $n_i = 1.5 \times 10^{10}/\text{cm}^3$, and assume that for intrinsic silicon $\mu_n = 1350\ \text{cm}^2/\text{V}\cdot\text{s}$ and $\mu_p = 480\ \text{cm}^2/\text{V}\cdot\text{s}$, and for the doped silicon $\mu_n = 1110\ \text{cm}^2/\text{V}\cdot\text{s}$ and $\mu_p = 400\ \text{cm}^2/\text{V}\cdot\text{s}$. (Note that doping results in reduced carrier mobilities.)

Solution

(a) For intrinsic silicon,

$$p = n = n_i = 1.5 \times 10^{10}/\text{cm}^3$$

Thus,

$$\rho = \frac{1}{q(p\mu_p + n\mu_n)}$$

$$\rho = \frac{1}{1.6 \times 10^{-19}(1.5 \times 10^{10} \times 480 + 1.5 \times 10^{10} \times 1350)}$$

$$= 2.28 \times 10^5\ \Omega\cdot\text{cm}$$

(b) For the p-type silicon

$$p_p \simeq N_A = 10^{16}/\text{cm}^3$$

$$n_p \simeq \frac{n_i^2}{N_A} = \frac{(1.5 \times 10^{10})^2}{10^{16}} = 2.25 \times 10^4/\text{cm}^3$$

Thus,

$$\rho = \frac{1}{q(p\mu_p + n\mu_n)}$$

$$= \frac{1}{1.6 \times 10^{-19}(10^{16} \times 400 + 2.25 \times 10^4 \times 1110)}$$

$$\simeq \frac{1}{1.6 \times 10^{-19} \times 10^{16} \times 400} = 1.56\ \Omega\cdot\text{cm}$$

Note that the resistivity of the p-type silicon is determined almost entirely by the doping concentration. Notice also that doping the silicon reduces its resistivity by a factor of about 10^4, a remarkable change.

EXERCISE

3.4 A uniform bar of n-type silicon of 2-μm length has a voltage of 1 V applied across it. If $N_D = 10^{16}/\text{cm}^3$ and $\mu_n = 1350\ \text{cm}^2/\text{V}\cdot\text{s}$, find (a) the electron drift velocity, (b) the time it takes an electron to cross the 2-μm length, (c) the drift-current density, and (d) the resistance and drift current in the case that the silicon bar has a cross-sectional area of 0.25 μm^2.

Ans. 6.75×10^6 cm/s; 30 ps; $1.08 \times 10^4\ \text{A/cm}^2$; 37.0 kΩ, 27 μA

3.3.2 Diffusion Current

Carrier diffusion occurs when the density of charge carriers in a piece of semiconductor is not uniform. For instance, if by some mechanism the concentration of, say, holes, is made higher in one part of a piece of silicon than in another, then holes will diffuse from the region of high concentration to the region of low concentration. The diffusion process is like what we would see if we dropped a few ink drops in a water-filled tank. The diffusion of charge carriers gives rise to a net flow of charge, or **diffusion current**.

As an example, consider the bar of silicon shown in Fig. 3.6(a): By some unspecified process, we have arranged to inject holes into its left side. This continuous hole injection gives rise to and maintains a hole **concentration profile** such as that shown in Fig. 3.6(b).

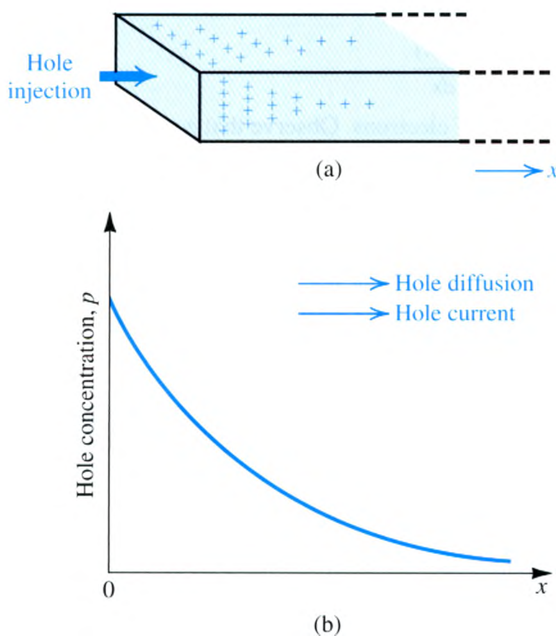


Figure 3.6 A bar of silicon **(a)** into which we inject holes, thus creating the hole concentration profile along the x axis, shown in **(b)**. The holes diffuse in the positive direction of x and give rise to a hole diffusion current in the same direction. Note that we are not showing the circuit to which the silicon bar is connected.

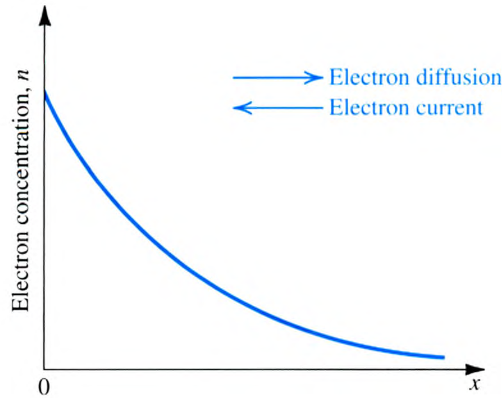


Figure 3.7 If we establish the electron concentration profile shown in a bar of silicon, electrons diffuse in the x direction, giving rise to an electron diffusion current in the negative-x direction.

This profile in turn causes holes to diffuse from left to right along the silicon bar, resulting in a hole current in the x direction. The magnitude of the current at any point is proportional to the slope of the concentration profile, or the **concentration gradient**, at that point,

$$J_{D,p} = -qD_p \frac{dp(x)}{dx} \tag{3.18}$$

where $J_{D,p}$ is the hole-current density (A/cm^2), q is the magnitude of electron charge, D_p is a constant called the **diffusion constant** or **diffusivity** of holes; and $p(x)$ is the hole concentration at point x. Note that the gradient (dp/dx) is negative, resulting in a positive current in the x direction, as should be expected.

In the case of electron diffusion resulting from an electron concentration gradient (see Fig. 3.7), a similar relationship applies, giving the electron-current density,

$$J_{D,n} = qD_n \frac{dn(x)}{dx} \tag{3.19}$$

where D_n is the diffusion constant or diffusivity of electrons. Observe that a negative (dn/dx) gives rise to a negative current, since the positive direction of current is that of the flow of positive net charge (and opposite to that of the flow of negative charge). In intrinsic silicon, typical values for the diffusion constants are $D_p = 12\ \text{cm}^2/\text{s}$ and $D_n = 35\ \text{cm}^2/\text{s}$.

At this point you may be wondering where the diffusion current in the silicon bar in Fig. 3.6(a) goes. A good question, as we are not showing how the right-side end of the bar is connected to the rest of the circuit. We will address this and related questions in detail in our discussion of the pn junction later in this chapter.

Example 3.4

Consider a bar of silicon with a hole concentration profile described by

$$p(x) = p_0 e^{-x/L_p}$$

Find the hole-current density at $x = 0$. Let $p_0 = 10^{16}/\text{cm}^3$, $L_p = 1\ \mu\text{m}$, and $D_p = 12\ \text{cm}^2/\text{s}$. If the cross-sectional area of the bar is $100\ \mu\text{m}^2$, find the current $I_{D,p}$.

Solution

Hole current will arise due to diffusion,

$$J_{D,p} = -qD_p \frac{dp(x)}{dx} = -qD_p \frac{d}{dx}\left[p_0 e^{-x/L_p}\right] = q\frac{D_p}{L_p} p_0 e^{-x/L_p}$$

Thus,

$$J_{D,p}(0) = q\frac{D_p}{L_p} p_0 = 1.6 \times 10^{-19} \times \frac{12}{1 \times 10^{-4}} \times 10^{16} = 192 \text{ A/cm}^2$$

The current $I_{D,p}$ can be found from

$$I_{D,p} = J_{D,p} \times A = 192 \times 100 \times 10^{-8} = 192\ \mu\text{A}$$

EXERCISE

3.5 The linear electron-concentration profile shown in Fig. E3.5 has been established in a piece of silicon. If $n_0 = 10^{17}/\text{cm}^3$ and $W = 0.5\ \mu\text{m}$, find the electron-current density in microamperes per micron squared ($\mu\text{A}/\mu\text{m}^2$). If a diffusion current of 1 mA is required, what must the cross-sectional area (in a direction perpendicular to the page) be? Recall that $D_n = 35\ \text{cm}^2/\text{s}$.

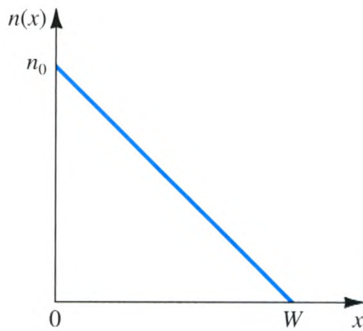


Figure E3.5

Ans. $112\ \mu\text{A}/\mu\text{m}^2$; $9\ \mu\text{m}^2$

3.3.3 Relationship between D and μ

A simple but powerful relationship relates the diffusion constant with the mobility,

$$\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = V_T \tag{3.20}$$

where $V_T = kT/q$. The parameter V_T is known as the **thermal voltage**. At room temperature, $T \simeq 300$ K and $V_T = 25.9$ mV. We will encounter V_T repeatedly throughout this book. The relationship in Eq. (3.20) is known as the **Einstein relationship**.

EXERCISE

3.6 Use the Einstein relationship to find D_n and D_p for intrinsic silicon using $\mu_n = 1350\ \text{cm}^2/\text{V}\cdot\text{s}$ and $\mu_p = 480\ \text{cm}^2/\text{V}\cdot\text{s}$.
Ans. 35 cm^2/s; 12.4 cm^2/s

3.4 The *pn* Junction

Having learned important semiconductor concepts, we are now ready to consider our first practical semiconductor structure: the *pn* junction. As mentioned previously, the *pn* junction implements the diode (Chapter 4) and plays the dominant role in the structure and operation of the bipolar junction transistor (BJT, Chapter 6). As well, understanding *pn* junctions is very important to the study of the MOSFET operation (Chapter 5).

3.4.1 Physical Structure

Figure 3.8 shows a simplified physical structure of the *pn* junction. It consists of a *p*-type semiconductor (e.g., silicon) brought into close contact with an *n*-type semiconductor material (also silicon). In actual practice, both the *p* and *n* regions are part of the same silicon crystal; that is, the *pn* junction is formed within a single silicon crystal by creating regions of different dopings (*p* and *n* regions). Appendix A provides a description of the fabrication process of integrated circuits including *pn* junctions. As indicated in Fig. 3.8, external wire connections are made to the *p* and *n* regions through metal contacts. If the *pn* junction is used as a diode, these constitute the diode terminals and we therefore label them "anode" and "cathode" in keeping with diode terminology.[3]

[3]The terminology is a carryover from vacuum-tube technology, used for making diodes and other electronic devices until the invention of the transistor in 1947. This event ushered in the era of solid-state electronics, which changed not only electronics, communications, and computers but indeed the world!

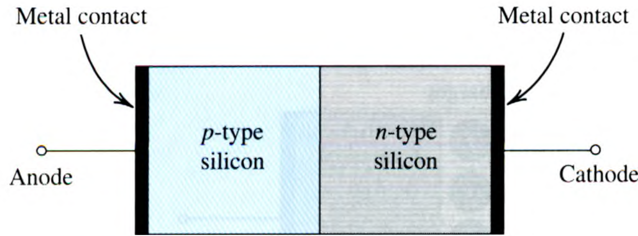


Figure 3.8 Simplified physical structure of the *pn* junction. (Actual geometries are given in Appendix A.) As the *pn* junction implements the junction diode, its terminals are labeled anode and cathode.

3.4.2 Operation with Open-Circuit Terminals

Figure 3.9 shows a *pn* junction under open-circuit conditions—that is, the external terminals are open. The "+" signs in the *p*-type material denote the majority holes. The charge of these holes is neutralized by an equal amount of bound negative charge associated with the acceptor atoms. For simplicity, these bound charges are not shown in the diagram. Also not shown are the minority electrons generated in the *p*-type material by thermal ionization.

In the *n*-type material the majority electrons are indicated by "–" signs. Here also, the bound positive charge, which neutralizes the charge of the majority electrons, is not shown in order to keep the diagram simple. The *n*-type material also contains minority holes generated by thermal ionization but not shown in the diagram.

The Diffusion Current I_D Because the concentration of holes is high in the *p* region and low in the *n* region, holes diffuse across the junction from the *p* side to the *n* side. Similarly, electrons diffuse across the junction from the *n* side to the *p* side. These two current components add together to form the diffusion current I_D, whose direction is from the *p* side to the *n* side, as indicated in Fig. 3.9.

The Depletion Region The holes that diffuse across the junction into the *n* region quickly recombine with some of the majority electrons present there and thus disappear from the scene. This recombination process results also in the disappearance of some free electrons from the *n*-type material. Thus some of the bound positive charge will no longer be neutralized by free electrons, and this charge is said to have been **uncovered**. Since recombination takes place close to the junction, there will be a region close to the junction that is *depleted of free electrons* and contains uncovered bound positive charge, as indicated in Fig. 3.9.

The electrons that diffuse across the junction into the *p* region quickly recombine with some of the majority holes there, and thus disappear from the scene. This results also in the disappearance of some majority holes, causing some of the bound negative charge to be uncovered (i.e., no longer neutralized by holes). Thus, in the *p* material close to the junction, there will be a region *depleted of holes* and containing uncovered bound negative charge, as indicated in Fig. 3.9.

From the above it follows that a **carrier-depletion region** will exist on both sides of the junction, with the *n* side of this region positively charged and the *p* side negatively charged. This carrier-depletion region—or, simply, **depletion region**—is also called the **space-charge region**. The charges on both sides of the depletion region cause an electric field E to be established across the region in the direction indicated in Fig. 3.9. Hence a potential difference results across the depletion region, with the *n* side at a positive voltage

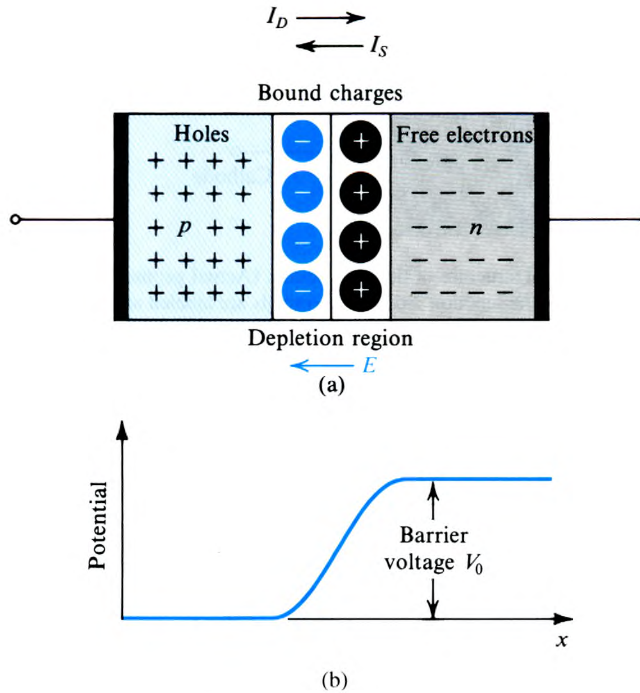


Figure 3.9 **(a)** The *pn* junction with no applied voltage (open-circuited terminals). **(b)** The potential distribution along an axis perpendicular to the junction.

relative to the *p* side, as shown in Fig. 3.9(b). Thus the resulting electric field opposes the diffusion of holes into the *n* region and electrons into the *p* region. In fact, the voltage drop across the depletion region acts as a **barrier** that has to be overcome for holes to diffuse into the *n* region and electrons to diffuse into the *p* region. The larger the barrier voltage, the smaller the number of carriers that will be able to overcome the barrier, and hence the lower the magnitude of diffusion current. Thus it is the appearance of the barrier voltage V_0 that limits the carrier diffusion process. It follows that the diffusion current I_D depends strongly on the voltage drop V_0 across the depletion region.

The Drift Current I_S and Equilibrium In addition to the current component I_D due to majority-carrier diffusion, a component due to minority-carrier drift exists across the junction. Specifically, some of the thermally generated holes in the *n* material move toward the junction and reach the edge of the depletion region. There, they experience the electric field in the depletion region, which sweeps them across that region into the *p* side. Similarly, some of the minority thermally generated electrons in the *p* material move to the edge of the depletion region and get swept by the electric field in the depletion region across that region into the *n* side. These two current components—electrons moved by drift from *p* to *n* and holes moved by drift from *n* to *p*—add together to form the drift current I_S, whose direction is from the *n* side to the *p* side of the junction, as indicated in Fig. 3.9. Since the current I_S is carried by thermally generated minority carriers, its value is strongly dependent on temperature; however, it is independent of the value of the depletion-layer voltage V_0. This is because the drift current is determined by the number of minority carriers that make it to the edge of the depletion region; any minority carriers that manage to get to the edge of the

depletion region will be swept across by E irrespective of the value of E or, correspondingly, of V_0.

Under open-circuit conditions (Fig. 3.9) no external current exists; thus the two opposite currents across the junction must be equal in magnitude:

$$I_D = I_S$$

This equilibrium condition[4] is maintained by the barrier voltage V_0. Thus, if for some reason I_D exceeds I_S, then more bound charge will be uncovered on both sides of the junction, the depletion layer will widen, and the voltage across it (V_0) will increase. This in turn causes I_D to decrease until equilibrium is achieved with $I_D = I_S$. On the other hand, if I_S exceeds I_D, then the amount of uncovered charge will decrease, the depletion layer will narrow, and the voltage across it (V_0) will decrease. This causes I_D to increase until equilibrium is achieved with $I_D = I_S$.

The Junction Built-in Voltage With no external voltage applied, the barrier voltage V_0 across the *pn* junction can be shown to be given by[5]

$$V_0 = V_T \ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{3.21}$$

where N_A and N_D are the doping concentrations of the p side and n side of the junction, respectively. Thus V_0 depends both on doping concentrations and on temperature since these impact the diffusion current, and thus the barrier voltage required to oppose it. The voltage V_0 is known as the **junction built-in voltage**. Typically, for silicon at room temperature, V_0 is in the range of 0.6 V to 0.9 V.

When the *pn* junction terminals are left open-circuited, the voltage measured between them will be zero. That is, the voltage V_0 across the depletion region *does not* appear between the junction terminals. This is because similar built-in voltages arise at the metal–semiconductor junctions at the terminals, which counter and exactly balance the barrier voltage. If this were not the case, we would be able to draw energy from the isolated *pn* junction, which would clearly violate the principle of conservation of energy.

Width of and Charge Stored in the Depletion Region Figure 3.10 provides further illustration of the situation in the *pn* junction in equilibrium. In Fig. 3.10(a) we show a junction in which $N_A > N_D$, a common situation in practice. The carrier concentrations on both sides of the junction are shown in Fig. 3.10(b). We have denoted the minority-carrier concentrations in both sides by n_{p0} and p_{n0}, with the additional subscript "0" signifying equilibrium (i.e., before external voltages are applied, as you will see in the next section). The depletion region extends in both the p and n materials and equal amounts of charge exist on both sides (Q_+ and Q_- in Fig. 3.10c). However, since we usually use unequal dopings N_A and N_D, as in Fig. 3.10, the width of the depletion layer will not be the same on the two sides. Rather, to uncover the same amount of charge, the depletion layer will extend deeper into the more lightly doped material. Specifically, if we denote the width of the

---

[4]In fact, in equilibrum the equality of drift and diffusion currents applies not just to the total currents but also to their individual components. That is, the hole drift current must equal the hole diffusion current and, similarly, the electron drift current must equal the electron diffusion current.

[5]The derivation of this formula and of a number of others in this chapter can be found in textbooks dealing with devices, such as that by Streetman and Bannerjee (see the reading list in Appendix I).

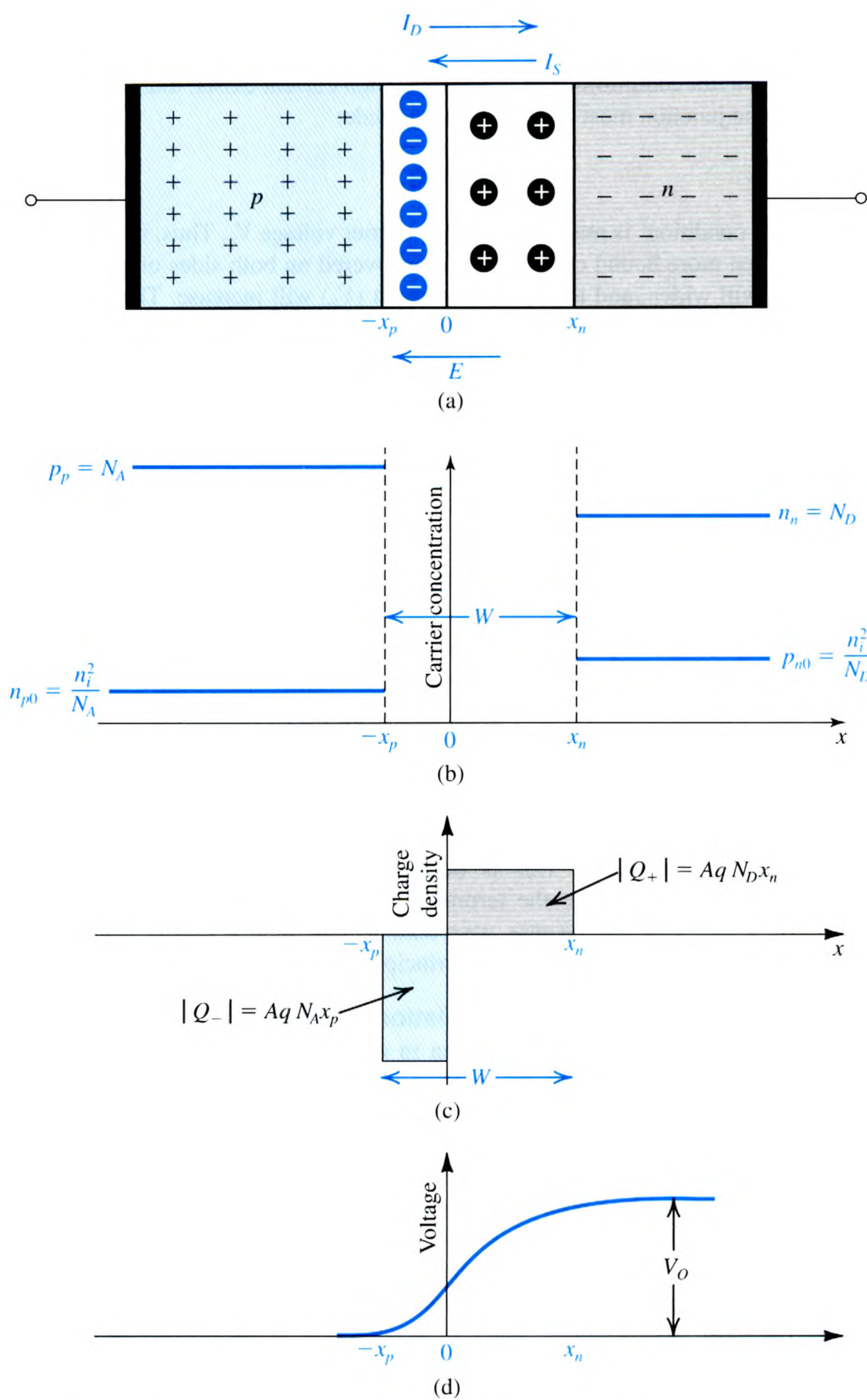


Figure 3.10 **(a)** A *pn* junction with the terminals open-circuited. **(b)** Carrier concentrations; note that $N_A > N_D$. **(c)** The charge stored in both sides of the depletion region; $Q_J = |Q_+| = |Q_-|$. **(d)** The built-in voltage V_0.

depletion region in the p side by x_p and in the n side by x_n, we can express the magnitude of the charge on the n side of the junction as

$$|Q_+| = qAx_nN_D \tag{3.22}$$

and that on the p side of the junction as

$$|Q_-| = qAx_pN_A \tag{3.23}$$

where A is the cross-sectional area of the junction in the plane perpendicular to the page. The charge equality condition can now be written as

$$qAx_nN_D = qAx_pN_A$$

which can be rearranged to yield

$$\frac{x_n}{x_p} = \frac{N_A}{N_D} \tag{3.24}$$

The width W of the depletion layer can be shown to be given by

$$W = x_n + x_p = \sqrt{\frac{2\epsilon_s}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)V_0} \tag{3.25}$$

where ϵ_s is the electrical permittivity of silicon $= 11.7\epsilon_0 = 11.7 \times 8.85 \times 10^{-14}$ F/cm $= 1.04 \times 10^{-12}$ F/cm. Typically W is in the range 0.1 μm to 1 μm. We can use Eqs. (3.24) and (3.25) to obtain x_n and x_p in terms of W as

$$x_n = W\frac{N_A}{N_A + N_D} \tag{3.26}$$

$$x_p = W\frac{N_D}{N_A + N_D} \tag{3.27}$$

We can express the charge stored on either side of the depletion region in terms of W by using Eqs. (3.22) and (3.26) to obtain

$$Q_J = |Q_+| = |Q_-|$$

$$Q_J = Aq\left(\frac{N_AN_D}{N_A + N_D}\right)W \tag{3.28}$$

Finally, we can substitute for W from Eq. (3.25) to obtain

$$Q_J = A\sqrt{2\epsilon_s q\left(\frac{N_AN_D}{N_A + N_D}\right)V_0} \tag{3.29}$$

These expressions for Q_J will prove useful in subsequent sections.

Usually, one side of the junction is much more heavily doped than the other, and the depletion region exists almost entirely on the more lightly doped side. For example, if $N_A \gg N_D$, then $W \approx x_n \gg x_p$. These are *one-sided junctions*.

Example 3.5

Consider a *pn* junction in equilibrium at room temperature ($T = 300$ K) with doping concentrations of $N_A = 10^{18}/\text{cm}^3$ and $N_D = 10^{16}/\text{cm}^3$ and a cross-sectional area $A = 10^{-4}$ cm^2. Calculate p_p, n_{p0}, n_n, p_{n0}, V_0, W, x_n, x_p, and Q_J. Use $n_i = 1.5 \times 10^{10}/\text{cm}^3$.

Solution

$$p_p \simeq N_A = 10^{18} \text{ cm}^{-3}$$

$$n_{p0} = \frac{n_i^2}{p_p} \simeq \frac{n_i^2}{N_A} = \frac{(1.5 \times 10^{10})^2}{10^{18}} = 2.25 \times 10^2 \text{ cm}^{-3}$$

$$n_n \simeq N_D = 10^{16} \text{ cm}^{-3}$$

$$p_{n0} = \frac{n_i^2}{n_n} \simeq \frac{n_i^2}{N_D} = \frac{(1.5 \times 10^{10})^2}{10^{16}} = 2.25 \times 10^4 \text{ cm}^{-3}$$

To find V_0 we use Eq. (3.21),

$$V_O = V_T \ln\left(\frac{N_A N_D}{n_i^2}\right)$$

where

$$V_T = \frac{kT}{q} = \frac{8.62 \times 10^{-5} \times 300 \text{ (eV)}}{1 \quad \text{(e)}} = 25.9 \times 10^{-3} \text{ V}$$

Thus,

$$V_0 = 25.9 \times 10^{-3} \ln\left(\frac{10^{18} \times 10^{16}}{2.25 \times 10^{20}}\right) = 0.814 \text{ V}$$

To determine W we use Eq. (3.25):

$$W = \sqrt{\frac{2 \times 1.04 \times 10^{-12}}{1.6 \times 10^{-19}}\left(\frac{1}{10^{18}} + \frac{1}{10^{16}}\right) \times 0.814} = 3.27 \times 10^{-5} \text{ cm} = 0.327 \text{ μm}$$

To determine x_n and x_p we use Eqs. (3.26) and (3.27), respectively:

$$x_n = W\frac{N_A}{N_A + N_D} = 0.327\frac{10^{18}}{10^{18} + 10^{16}} = 0.324 \text{ μm}$$

$$x_p = W\frac{N_D}{N_A + N_D} = 0.327\frac{10^{16}}{10^{18} + 10^{16}} = 0.003 \text{ μm}$$

Finally, to determine the charge stored on either side of the depletion region, we use Eq. (3.28):

$$Q_J = 10^{-4} \times 1.6 \times 10^{-19}\left(\frac{10^{18} \times 10^{16}}{10^{18} + 10^{16}}\right) \times 0.327 \times 10^{-4} = 5.18 \times 10^{-12} \text{ C} = 5.18 \text{ pC}$$

EXERCISES

3.7 Show that

$$V_0 = \frac{1}{2}\left(\frac{q}{\epsilon_s}\right)\left(\frac{N_A N_D}{N_A + N_D}\right)W^2$$

3.8 Show that for a one-sided *pn* junction in which the *p* side is much more heavily doped than the *n* side (i.e., $N_A \gg N_D$), referred to as a p^+n diode, Eqs. (3.25), (3.26), (3.27), (3.28), and (3.29) can be simplified as follows:

$$W \simeq \sqrt{\frac{2\epsilon_s}{qN_D}V_0} \quad (3.25')$$

$$x_n \simeq W \quad (3.26')$$

$$x_p \simeq W/(N_A/N_D) \quad (3.27')$$

$$Q_J \simeq AqN_D W \quad (3.28')$$

$$Q_J \simeq A\sqrt{2\epsilon_s qN_D V_0} \quad (3.29')$$

3.9 If in the fabrication of the *pn* junction in Example 3.5, it is required to increase the minority-carrier concentration in the *n* region by a factor of 2, what must be done?
Ans. Lower N_D by a factor of 2.

3.5 The *pn* Junction with an Applied Voltage

Having studied the open-circuited *pn* junction in detail, we are now ready to apply a dc voltage between its two terminals to find its electrical conduction properties. If we apply a voltage so that the *p* side is made more positive than the *n* side, it is referred to as a forward-bias[6] voltage. Conversely, if we make the *n* side more positive than the *p* side, it is said to be a reverse-bias voltage. As you will see, the *pn* junction exhibits vastly different conduction properties in its forward and reverse directions.

We will begin with a simple qualitative description in Section 3.5.1 and then consider an analytical description of the *i*–*v* characteristic of the junction in Section 3.5.2.

3.5.1 Qualitative Description of Junction Operation

Figure 3.11 shows the *pn* junction under three different conditions: (a) the open-circuit or equilibrium condition studied in the previous section; (b) the reverse-bias condition, where a dc voltage V_R is applied; and (c) the forward-bias condition, where a dc voltage V_F is applied.

[6]For the time being, we take the term *bias* to refer simply to the application of a dc voltage. We will see in later chapters that it has a deeper meaning in the design of electronic circuits.

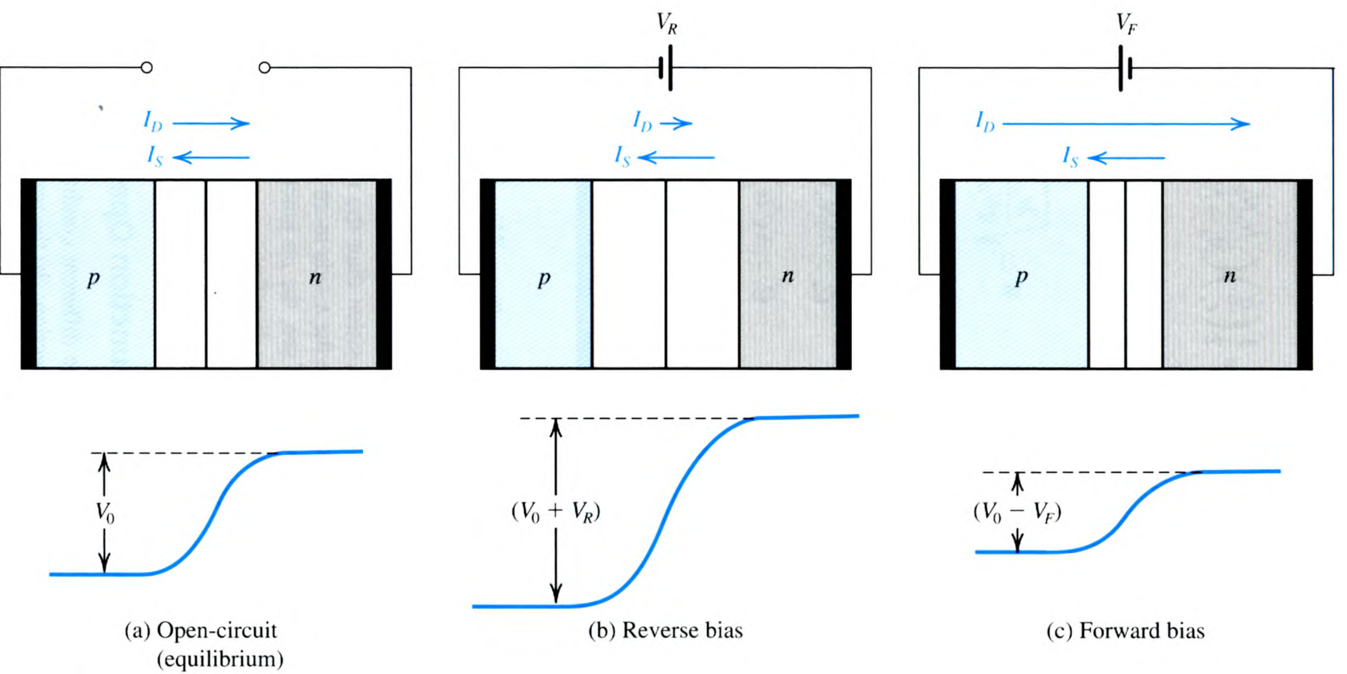


Figure 3.11 The *pn* junction in: **(a)** equilibrium; **(b)** reverse bias; **(c)** forward bias.

Notice that in the open-circuit case, a barrier voltage V_0 develops, making n more positive than p, and limiting the diffusion current I_D to a value exactly equal to the drift current I_S. This results in a zero current at the junction terminals, as should be the case, since the terminals are open-circuited. Also, as mentioned previously, the barrier voltage V_0, though it establishes the current equilibrium across the junction, does *not* in fact appear between the junction terminals because opposing built-in voltages arise at the terminals' metal contacts.

Consider now the reverse-bias case in (b). The externally applied reverse-bias voltage V_R adds to the barrier voltage, thus increasing the effective barrier voltage to $(V_0 + V_R)$ as shown. This reduces the number of holes that diffuse into the n region and the number of electrons that diffuse into the p region. As a result, the diffusion current I_D is dramatically reduced. As you will see shortly, a reverse-bias voltage of a volt or so is sufficient to cause $I_D \simeq 0$, and the current across the junction and through the external circuit will be equal to I_S. Recalling that I_S is the current due to the drift across the depletion region of the thermally generated minority carriers, we expect I_S to be very small and to be strongly dependent on temperature. We will show this to be the case very shortly. We conclude that in the reverse direction, the *pn* junction conducts a very small and almost constant current equal to I_S.

Before leaving the reverse-bias case, observe that the increase in barrier voltage will be accompanied by a corresponding increase in the stored uncovered charge on both sides of the depletion region. This in turn means a wider depletion region, needed to uncover the additional charge required to support the larger barrier voltage $(V_0 + V_R)$. Analytically, we can obtain these results by extending the results of the equilibrium case. Thus we obtain the width of the depletion region by replacing V_0 in Eq. (3.25) by $(V_0 + V_R)$,

$$W = x_n + x_p = \sqrt{\frac{2\epsilon_s}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)(V_0 + V_R)} \tag{3.30}$$

and the magnitude of the charge stored on either side of the depletion region by replacing V_0 in Eq. (3.29) by $(V_0 + V_R)$,

$$Q_J = A\sqrt{2\epsilon_s q\left(\frac{N_A N_D}{N_A + N_D}\right)(V_0 + V_R)} \tag{3.31}$$

We next consider the forward-bias case shown in Fig. 3.11(c). Here the applied voltage V_F is in the direction that subtracts from the built-in voltage V_0, resulting in a reduced barrier voltage $(V_0 - V_F)$ across the depletion region. This reduced barrier voltage will be accompanied by reduced depletion-region charge and correspondingly narrower depletion-region width W. Most importantly, the lowering of the barrier voltage will enable more holes to diffuse from p to n and more electrons to diffuse from n to p. Thus the diffusion current I_D increases substantially and can become many orders of magnitude larger than the drift current I_S. The current I in the external circuit is, of course, the difference between I_D and I_S,

$$I = I_D - I_S$$

and it flows in the forward direction of the junction, from p to n. We thus conclude that the *pn* junction can conduct a substantial current in the forward-bias region and that current is mostly a diffusion current whose value is determined by the forward-bias voltage V_F.

3.5.2 The Current–Voltage Relationship of the Junction

We are ready to find an analytical expression that describes the current–voltage relationship of the *pn* junction. We will consider a junction operating with a forward applied voltage

V and derive an expression for the current I that flows in the forward direction (from p to n). However, our derivation is general and will yield the reverse current when the applied voltage V is made negative.

From the qualitative description above we know that a forward-bias voltage V subtracts from the built-in voltage V_0, thus resulting in a lower barrier voltage $(V_0 - V)$. The lowered barrier in turn makes it possible for a greater number of holes to overcome the barrier and diffuse into the n region. We can make a similar statement about electrons from the n region diffusing into the p region.

Let us now consider the holes injected into the n region. The concentration of holes in the n region at the edge of the depletion region will increase considerably. In fact, an important result from device physics shows that the steady-state concentration at the edge of the depletion region will be

$$p_n(x_n) = p_{n0}e^{V/V_T} \tag{3.32}$$

That is, the concentration of the minority holes increases from the equilibrium value of p_{n0} (see Fig. 3.10) to the much larger value determined by the value of V, given by Eq. (3.32).

We describe this situation as follows: The forward-bias voltage V results in an **excess concentration** of minority holes at $x = x_n$, given by

$$\begin{aligned}\text{Excess concentration} &= p_{n0}e^{V/V_T} - p_{n0} \\ &= p_{n0}\left(e^{V/V_T} - 1\right)\end{aligned} \tag{3.33}$$

The increase in minority-carrier concentration in Eqs. (3.32) and (3.33) occurs at the edge of the depletion region $(x = x_n)$. As the injected holes diffuse into the n material, some will recombine with the majority electrons and disappear. Thus, the excess hole concentration will decay exponentially with distance. As a result, the total hole concentration in the n material will be given by

$$p_n(x) = p_{n0} + (\text{Excess concentration})e^{-(x-x_n)/L_p}$$

Substituting for the "Excess concentration" from Eq. (3.33) gives

$$p_n(x) = p_{n0} + p_{n0}\left(e^{V/V_T} - 1\right)e^{-(x-x_n)/L_p} \tag{3.34}$$

The exponential decay is characterized by the constant L_p, which is called the **diffusion length** of holes in the n material. The smaller the value of L_p, the faster the injected holes will recombine with the majority electrons, resulting in a steeper decay of minority-carrier concentration.

Figure 3.12 shows the steady-state minority-carrier concentration profiles on both sides of a pn junction in which $N_A \gg N_D$. Let's stay a little longer with the diffusion of holes into the n region. Note that the shaded region under the exponential represents the excess minority carriers (holes). From our study of diffusion in Section 3.3, we know that the establishment of a carrier concentration profile such as that in Fig. 3.12 is essential to support a steady-state diffusion current. In fact, we can now find the value of the hole–diffusion current density by applying Eq. (3.18),

$$J_p(x) = -qD_p\frac{dp_n(x)}{dx}$$

Substituting for $p_n(x)$ from Eq. (3.34) gives

$$J_p(x) = q\left(\frac{D_p}{L_p}\right)p_{n0}\left(e^{V/V_T} - 1\right)e^{-(x-x_n)/L_p} \tag{3.35}$$

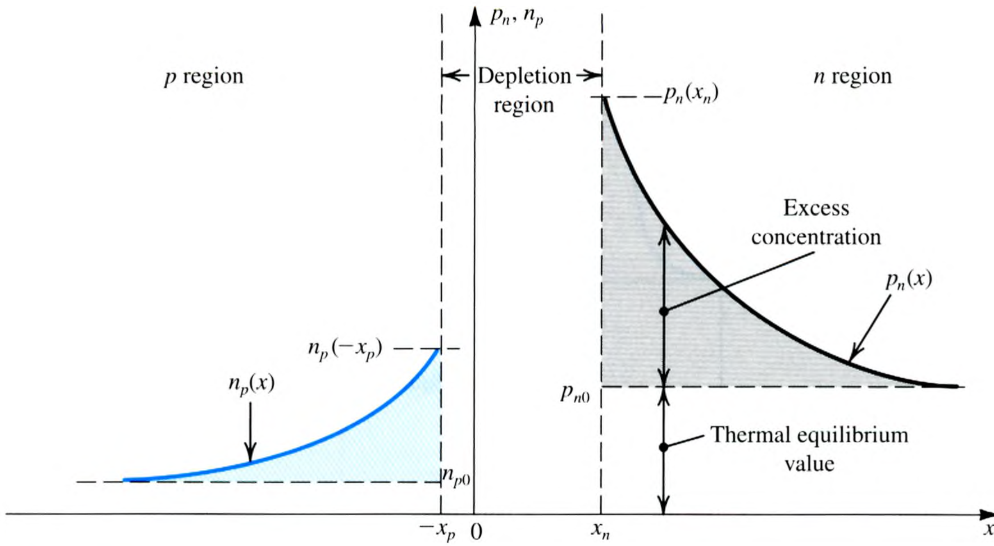


Figure 3.12 Minority-carrier distribution in a forward-biased *pn* junction. We assume that the *p* region is more heavily doped than the *n* region; $N_A \gg N_D$.

As expected, $J_p(x)$ is highest at $x = x_n$,

$$J_p(x_n) = q\left(\frac{D_p}{L_p}\right)p_{n0}(e^{V/V_T} - 1) \tag{3.36}$$

and decays exponentially for $x > x_n$, as the minority holes recombine with the majority electrons. This recombination, however, means that the majority electrons will have to be replenished by a current that injects electrons from the external circuit into the *n* region of the junction. This latter current component has the same direction as the hole current (because electrons moving from right to left give rise to current in the direction from left to right). It follows that as $J_p(x)$ decreases, the electron current component increases by exactly the same amount, making the total current in the *n* material constant at the value given by Eq. (3.36).

An exactly parallel development can be applied to the electrons that are injected from the *n* to the *p* region, resulting in an electron diffusion current given by a simple adaptation of Eq. (3.36),

$$J_n(-x_p) = q\left(\frac{D_n}{L_n}\right)n_{p0}(e^{V/V_T} - 1) \tag{3.37}$$

Now, although the currents in Eqs. (3.36) and (3.37) are found at the two edges of the depletion region, their values do not change in the depletion region. Thus we can drop the location descriptors (x_n), $(-x_p)$, add the two current densities, and multiply by the junction area A to obtain the total current I as

$$I = A(J_p + J_n)$$

$$I = Aq\left(\frac{D_p}{L_p}p_{n0} + \frac{D_n}{L_n}n_{p0}\right)(e^{V/V_T} - 1)$$

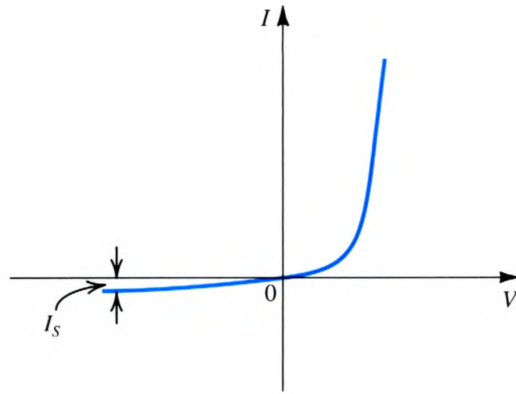


Figure 3.13 The *pn* junction *I–V* characteristic.

Substituting for $p_{n0} = n_i^2/N_D$ and for $n_{p0} = n_i^2/N_A$ gives

$$I = Aqn_i^2\left(\frac{D_p}{L_pN_D} + \frac{D_n}{L_nN_A}\right)\left(e^{V/V_T} - 1\right) \tag{3.38}$$

From this equation we note that for a negative *V* (reverse bias) with a magnitude of a few times V_T (25.9 mV at room temperature), the exponential term becomes essentially zero, and the current across the junction becomes negative and constant. From our qualitative description in Section 3.5.1, we know that this current must be I_S. Thus,

$$I = I_S\left(e^{V/V_T} - 1\right) \tag{3.39}$$

where

$$I_S = Aqn_i^2\left(\frac{D_p}{L_pN_D} + \frac{D_n}{L_nN_A}\right) \tag{3.40}$$

Figure 3.13 shows the *I–V* characteristic of the *pn* junction (Eq. 3.39). Observe that in the reverse direction the current saturates at a value equal to $-I_S$. For this reason, I_S is given the name **saturation current**. From Eq. (3.40) we see that I_S is directly proportional to the cross-sectional area *A* of the junction. Thus, another name for I_S, one we prefer to use in this book, is the junction **scale current**. Typical values for I_S, for junctions of various areas, range from 10^{-18} A to 10^{-12} A.

Besides being proportional to the junction area *A*, the expression for I_S in Eq. (3.40) indicates that I_S is proportional to n_i^2, which is a very strong function of temperature (see Eq. 3.2).

Example 3.6

For the *pn* junction considered in Example 3.5 for which $N_A = 10^{18}/\text{cm}^3$, $N_D = 10^{16}/\text{cm}^3$, $A = 10^{-4}\text{cm}^2$, and $n_i = 1.5 \times 10^{10}/\text{cm}^3$, let $L_p = 5$ μm, $L_n = 10$ μm, D_p (in the *n* region) = 10 cm²/V·s, and D_n (in the *p* region) = 18 cm²/V·s. The *pn* junction is forward biased and conducting a current *I* = 0.1 mA.

Calculate: (a) I_S; (b) the forward-bias voltage V; and (c) the component of the current I due to hole injection and that due to electron injection across the junction.

Solution

(a) Using Eq. (3.40), we find I_S as

$$I_S = 10^{-4} \times 1.6 \times 10^{-19} \times \left(1.5 \times 10^{10}\right)^2 \times \left(\frac{10}{5 \times 10^{-4} \times 10^{16}} + \frac{18}{10 \times 10^{-4} \times 10^{18}}\right) = 7.3 \times 10^{-15}\,\text{A}$$

(b) In the forward direction,

$$I = I_S\left(e^{V/V_T} - 1\right) \simeq I_S e^{V/V_T}$$

Thus,

$$V = V_T \ln\left(\frac{I}{I_S}\right)$$

For $I = 0.1$ mA,

$$V = 25.9 \times 10^{-3} \ln\left(\frac{0.1 \times 10^{-3}}{7.3 \times 10^{-15}}\right) = 0.605\ \text{V}$$

(c) The hole-injection component of I can be found using Eq. (3.36)

$$I_p = Aq\frac{D_p}{L_p}p_{n0}\left(e^{V/V_T} - 1\right) = Aq\frac{D_p}{L_p}\frac{n_i^2}{N_D}\left(e^{V/V_T} - 1\right)$$

Similarly, I_n can be found using Eq. (3.38),

$$I_n = Aq\frac{D_n}{L_n}\frac{n_i^2}{N_A}\left(e^{V/V_T} - 1\right)$$

Thus,

$$\frac{I_p}{I_n} = \left(\frac{D_p}{D_n}\right)\left(\frac{L_n}{L_p}\right)\left(\frac{N_A}{N_D}\right)$$

For our case,

$$\frac{I_p}{I_n} = \frac{10}{18} \times \frac{10}{5} \times \frac{10^{18}}{10^{16}} = 1.11 \times 10^2 = 111$$

Thus most of the current is conducted by holes injected into the n region. Specifically,

$$I_p = \frac{111}{112} \times 0.1 = 0.0991\ \text{mA}$$

$$I_n = \frac{1}{112} \times 0.1 = 0.0009\ \text{mA}$$

This stands to reason, since the p material has a doping concentration 100 times that of the n material.

EXERCISES

3.10 Show that if $N_A \gg N_D$,

$$I_S \simeq Aqn_i^2 \frac{D_p}{L_p N_D}$$

3.11 For the *pn* junction in Example 3.6, find the value of I_S and that of the current I at $V = 0.605$ V (same voltage found in Example 3.6 at a current $I = 0.1$ mA) if N_D is reduced by a factor of 2.
Ans. 1.46×10^{-14} A; 0.2 mA

3.12 For the *pn* junction considered in Examples 3.5 and 3.6, find the width of the depletion region W corresponding to the forward-bias voltage found in Example 3.6. (*Hint*: Use the formula in Eq. (3.30) with V_R replaced with $-V_F$.)
Ans. 0.166 μm

3.13 For the *pn* junction considered in Examples 3.5 and 3.6, find the width of the depletion region W and the charge stored in the depletion region Q_J when a 2-V reverse bias is applied. Also find the value of the reverse current I.
Ans. 0.608 μm; 9.63 pC; 7.3×10^{-15} A

3.5.3 Reverse Breakdown

The description of the operation of the *pn* junction in the reverse direction, and the $I-V$ relationship of the junction in Eq. (3.39), indicate that at a reverse-bias voltage $-V$, with $V \gg V_T$, the reverse current that flows across the junction is approximately equal to I_S and thus is very small. However, as the magnitude of the reverse-bias voltage V is increased, a value is reached at which a very large reverse current flows as shown in Fig. 3.14. Observe that as V reaches the value V_{BR}, the dramatic increase in reverse current is accompanied by a very small increase in the reverse voltage; that is, the reverse voltage across the junction remains very close to the value V_{BR}. The phenomenon is known as **junction breakdown**. It is not a destructive phenomenon. That is, the *pn* junction can be repeatedly operated in the breakdown region without a permanent effect on its characteristics. This, however, is predicated on the assumption that the magnitude of the reverse-breakdown current is limited by the external circuit to a "safe" value. The "safe" value is one that results in the limitation of the power dissipated in the junction to a safe, allowable level.

There are two possible mechanisms for *pn* junction breakdown: the **zener effect** and the **avalanche effect**. If a *pn* junction breaks down with a breakdown voltage $V_{BR} < 5$ V, the breakdown mechanism is usually the zener effect. Avalanche breakdown occurs when V_{BR} is greater than approximately 8 V. For junctions that break down between 5 V and 8 V, the breakdown mechanism can be either the zener or the avalanche effect or a combination of the two.

Zener breakdown occurs when the electric field in the depletion layer increases to the point of breaking covalent bonds and generating electron–hole pairs. The electrons generated in this way will be swept by the electric field into the *n* side and the holes into the *p* side.

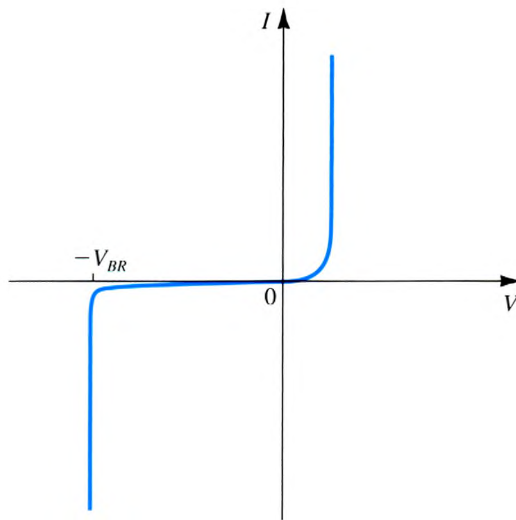


Figure 3.14 The I–V characteristic of the *pn* junction showing the rapid increase in reverse current in the breakdown region.

Thus these electrons and holes constitute a reverse current across the junction. Once the zener effect starts, a large number of carriers can be generated, with a negligible increase in the junction voltage. Thus the reverse current in the breakdown region will be large and its value must be determined by the external circuit, while the reverse voltage appearing between the diode terminals will remain close to the specified breakdown voltage V_{BR}, often denoted V_Z[7] in this case.

The other breakdown mechanism, avalanche breakdown, occurs when the minority carriers that cross the depletion region under the influence of the electric field gain sufficient kinetic energy to be able to break covalent bonds in atoms with which they collide. The carriers liberated by this process may have sufficiently high energy to be able to cause other carriers to be liberated in another ionizing collision. This process keeps repeating in the fashion of an avalanche, with the result that many carriers are created that are able to support any value of reverse current, as determined by the external circuit, with a negligible change in the voltage drop across the junction.

As will be seen in Chapter 4, some *pn* junction diodes are fabricated to operate specifically in the zener breakdown region, where use is made of the nearly constant voltage V_Z.

3.6 Capacitive Effects in the *pn* Junction

There are two charge-storage mechanisms in the *pn* junction. One is associated with the charge stored in the depletion region, and the other is associated with the minority-carrier charge stored in the *n* and *p* materials as a result of the concentration profiles established by carrier injection. While the first is easier to see when the *pn* junction is reverse biased, the second is in effect, and typically predominating, only when the junction is forward biased.

[7]Note that the subscript Z in V_Z denotes *zener*, named after an early worker in the area. We will use V_Z to denote the breakdown voltage when the breakdown mechanism is the zener effect.

3.6.1 Depletion or Junction Capacitance

When a *pn* junction is reverse biased with a voltage V_R, the charge stored on either side of the depletion region is given by Eq. (3.31),

$$Q_J = A\sqrt{2\epsilon_s q \frac{N_A N_D}{N_A + N_D}(V_0 + V_R)}$$

Thus, for a given *pn* junction,

$$Q_J = \alpha\sqrt{V_0 + V_R} \tag{3.41}$$

where α is given by

$$\alpha = A\sqrt{2\epsilon_s q \frac{N_A N_D}{N_A + N_D}} \tag{3.42}$$

In an ideal capacitor, the stored charge is linearly proportional to the applied voltage. In a *pn* junction, Eq. (3.41) tells us Q_J is nonlinearly related to V_R, as shown in Fig. 3.15. This nonlinear relationship makes it difficult to define a capacitance that accounts for the need to change Q_J whenever V_R is changed. We can, however, assume that the junction is operating at a point such as Q, as indicated in Fig. 3.15, and define a capacitance C_j that relates the change in the charge Q_J to a change in the voltage V_R,

$$C_j = \left.\frac{dQ_J}{dV_R}\right|_{V_R = V_Q} \tag{3.43}$$

This incremental-capacitance approach turns out to be quite useful in electronic circuit design, as we shall see throughout this book.

Using Eq. (3.43) together with Eq. (3.41) yields

$$C_j = \frac{\alpha}{2\sqrt{V_0 + V_R}} \tag{3.44}$$

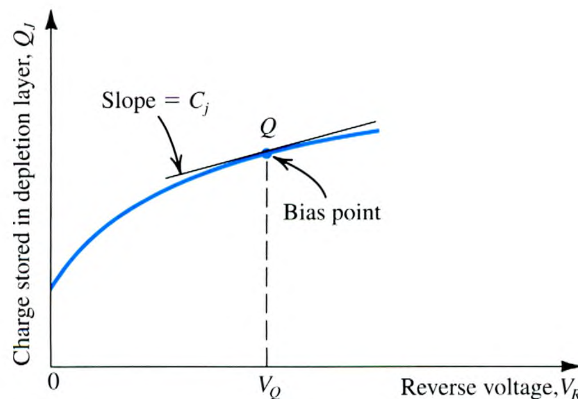


Figure 3.15 The charge stored on either side of the depletion layer as a function of the reverse voltage V_R.

The value of C_j at zero reverse bias can be obtained from Eqs. (3.42) and (3.44) as

$$C_{j0} = \frac{\alpha}{2\sqrt{V_0}} = A\sqrt{\left(\frac{\epsilon_s q}{2}\right)\left(\frac{N_A N_D}{N_A + N_D}\right)\left(\frac{1}{V_0}\right)} \tag{3.45}$$

which enables us to express C_j as

$$C_j = \frac{C_{j0}}{\sqrt{1 + \frac{V_R}{V_0}}} \tag{3.46}$$

Before leaving the subject of depletion-region or junction capacitance we point out that in the *pn* junction we have been studying, the doping concentration is made to change abruptly at the junction boundary. Such a junction is known as an **abrupt junction**. There is another type of *pn* junction in which the carrier concentration is made to change gradually from one side of the junction to the other. To allow for such a **graded junction**, the formula for the junction capacitance (Eq. 3.46) can be written in the more general form

$$C_j = \frac{C_{j0}}{\left(1 + \frac{V_R}{V_0}\right)^m} \tag{3.47}$$

where m is a constant called the **grading coefficient**, whose value ranges from 1/3 to 1/2 depending on the manner in which the concentration changes from the p to the n side.

EXERCISE

3.14 For the *pn* junction considered in Examples 3.5 and 3.6, find C_{j0} and C_j at $V_R = 2$ V. Recall that $V_0 = 0.814$ V, $N_A = 10^{18}/\text{cm}^3$, $N_D = 10^{16}/\text{cm}^3$, and $A = 10^{-4}$ cm^2.
Ans. 3.2 pF; 1.7 pF

3.6.2 Diffusion Capacitance

Consider a forward-biased *pn* junction. In steady state, minority-carrier distributions in the p and n materials are established, as shown in Fig. 3.12. Thus a certain amount of excess minority-carrier charge is stored in each of the p and n bulk regions (outside the depletion region). If the terminal voltage V changes, this charge will have to change before a new steady state is achieved. This charge-storage phenomenon gives rise to another capacitive effect, distinctly different from that due to charge storage in the depletion region.

To calculate the excess minority-carrier charge, refer to Fig. 3.12. The excess hole charge stored in the n region can be found from the shaded area under the exponential as

follows:[8]

$$Q_p = Aq \times \text{shaded area under the } p_n(x) \text{curve}$$
$$= Aq\left[p_n(x_n) - p_{n0}\right]L_p$$

Substituting for $p_n(x_n)$ from Eq. (3.32) and using Eq. (3.36) enables us to express Q_p as

$$Q_p = \frac{L_p^2}{D_p} I_p \tag{3.48}$$

The factor $\left(L_p^2/D_p\right)$ that relates Q_p to I_p is a useful device parameter that has the dimension of time (s) and is denoted τ_p

$$\tau_p = \frac{L_p^2}{D_p} \tag{3.49}$$

Thus,

$$Q_p = \tau_p I_p \tag{3.50}$$

The time constant τ_p is known as the excess **minority-carrier (hole) lifetime**. *It is the average time it takes for a hole injected into the n region to recombine with a majority electron*. This definition of τ_p implies that the entire charge Q_p disappears and has to be replenished every τ_p seconds. The current that accomplishes the replenishing is $I_p = Q_p/\tau_p$. This is an alternate derivation for Eq. (3.50).

A relationship similar to that in Eq. (3.50) can be developed for the electron charge stored in the p region,

$$Q_n = \tau_n I_n \tag{3.51}$$

where τ_n is the electron lifetime in the p region. The total excess minority-carrier charge can be obtained by adding together Q_p and Q_n,

$$Q = \tau_p I_p + \tau_n I_n \tag{3.52}$$

This charge can be expressed in terms of the diode current $I = I_p + I_n$ as

$$Q = \tau_T I \tag{3.53}$$

where τ_T is called the **mean transit time** of the junction. Obviously, τ_T is related to τ_p and τ_n. Furthermore, for most practical devices, one side of the junction is much more heavily doped than the other. For instance, if $N_A \gg N_D$, one can show that $I_p \gg I_n$, $I \simeq I_p$, $Q_p \gg Q_n$, $Q \simeq Q_p$, and thus $\tau_T \simeq \tau_p$.

For small changes around a bias point, we can define an **incremental diffusion capacitance** C_d as

$$C_d = \frac{dQ}{dV} \tag{3.54}$$

[8]Recall that the area under an exponential curve $Ae^{-x/B}$ from $x = 0$ to ∞ is equal to AB.

Using Eqs. (3.53) and (3.39) approximating $I \approxeq I_S e^{V/V_T}$, we can show that

$$C_d = \left(\frac{\tau_T}{V_T}\right) I \tag{3.55}$$

where I is the forward-bias current. Note that C_d is directly proportional to the forward current I and thus is negligibly small when the diode is reverse biased. Also note that to keep C_d small, the transit time τ_T must be made small, an important requirement for a *pn* junction intended for high-speed or high-frequency operation.

EXERCISES

3.15 Use the definition of C_d in Eq. (3.54) to derive the expression in Eq. (3.55) by means of Eqs. (3.53) and (3.39).

3.16 For the *pn* junction considered in Examples 3.5 and 3.6 for which $D_p = 10\ \text{cm}^2/\text{V}\cdot\text{s}$, and $L_p = 5\ \mu\text{m}$, find τ_p and C_d at a forward-bias current of 0.1 mA. Recall that for this junction, $I_p \simeq I$.
Ans. 25 ns; 96.5 pF

Summary

- Today's microelectronics technology is almost entirely based on the semiconductor material silicon. If a circuit is to be fabricated as a monolithic integrated circuit (IC) it is made using a single silicon crystal, no matter how large the circuit is (modern chips can contain billions of transistors).
- In a crystal of intrinsic or pure silicon, the atoms are held in position by covalent bonds. At very low temperatures, all the bonds are intact, and no charge carriers are available to conduct electrical current.
- At room temperature, thermal energy causes some of the covalent bonds to break, thus generating free electrons and holes that become available for current conduction.
- Current in semiconductors is carried by free electrons and holes. Their numbers are equal and relatively small in intrinsic silicon. Thus, intrinsic silicon is an insulator.
- The conductivity of silicon can be increased dramatically by introducing small amounts of appropriate impurity materials into the silicon crystal in a process called doping.
- There are two kinds of doped semiconductor: *n*-type, in which electrons are abundant, and *p*-type, in which holes are abundant.
- There are two mechanisms for the transport of charge carriers in semiconductors: drift and diffusion.
- Carrier drift results when an electric field E is applied across a piece of silicon. The electric field accelerates the holes in the direction of E and the electrons in the direction opposite to E. These two current components add together to produce a drift current in the direction of E.
- Carrier diffusion occurs when the concentration of charge carriers is made higher in one part of the silicon crystal than in other parts. To establish a steady-state diffusion current, a carrier concentration gradient must be maintained in the silicon crystal.
- A basic semiconductor structure is the *pn* junction. It is fabricated in a silicon crystal by creating a *p* region in close proximity to an *n* region. The *pn* junction is a diode and plays a dominant role in the structure and operation of transistors.
- When the terminals of the *pn* junction are left open, no current flows externally. However, two equal and opposite currents, I_D and I_S, flow across the junction, and equilibrium is maintained by a built-in voltage V_0 that develops across the junction, with the *n* side positive

relative to the p side. Note, however, that the voltage across an open junction is 0 V, since V_0 is canceled by potentials appearing at the metal-to-semiconductor connection interfaces.

- The voltage V_0 appears across the depletion region, which extends on both sides of the junction.
- The diffusion current I_D is carried by holes diffusing from p to n and electrons diffusing from n to p. I_D flows from p to n, which is the forward direction of the junction. Its value depends on V_0.
- The drift current I_S is carried by thermally generated minority electrons in the p material that are swept across the depletion layer into the n side, and by thermally generated minority holes in the n side that are swept across the depletion region into the p side. I_S flows from n to p, in the reverse direction of the junction, and its value is a strong function of temperature but independent of V_0.
- Forward biasing the pn junction (i.e. applying an external voltage V that makes p more positive than n) reduces the barrier voltage to $V_0 - V$ and produces an exponential increase in I_D while I_S remains unchanged. The net result is a substantial current $I = I_D - I_S$ that flows across the junction and through the external circuit.
- Applying a negative V reverse biases the junction and increases the barrier voltage, with the result that I_D is reduced to almost zero and the net current across the junction becomes the very small reverse current I_S.
- If the reverse voltage is increased in magnitude to a value V_{BR} specific to the particular junction, the junction breaks down, and a large reverse current flows. The value of the reverse current must be limited by the external circuit.
- Whenever the voltage across a pn junction is changed, some time has to pass before steady state is reached. This is due to the charge-storage effects in the junction, which are modeled by two capacitances: the junction capacitance C_j and the diffusion capacitance C_d.
- Table 3.1 (in the Summary Tables supplement found online at www.oup.com/he/sedra-smith8e.) provides a summary of pertinent relationships and the values of physical constants.

PROBLEMS

If in the following problems you require values of particular parameters or physical constants that are not stated, please consult Table 3.1 in the Summary Tables supplement found online at www.oup.com/he/sedra-smith8e.

Section 3.1: Intrinsic Semiconductors

3.1 Find values of the intrinsic carrier concentration n_i for silicon at −55°C, 0°C, 20°C, 75°C, and 125°C. At each temperature, what fraction of the atoms is ionized? Recall that a silicon crystal has approximately 5×10^{22} atoms/cm^3.

3.2 Calculate the value of n_i for gallium arsenide (GaAs) at $T = 300$ K. The constant $B = 3.56 \times 10^{14}$ cm^{-3}K$^{-3/2}$ and the bandgap voltage $E_g = 1.42$ eV. Compare with that of silicon at the same temperature.

3.3 Find the total number of free electrons in a circular wafer of single crystal silicon that is 300 μm thick and 15 cm in diameter at $T = 300$ K.

Section 3.2: Doped Semiconductors

3.4 For a p-type silicon in which the dopant concentration $N_A = 2 \times 10^{18}$/cm^3, find the hole and electron concentrations at $T = 300$ K.

3.5 For a silicon crystal doped with phosphorus, what must N_D be if at $T = 300$ K the hole concentration drops below the intrinsic level by a factor of 10^8?

3.6 In a boron-doped silicon layer with impurity concentration of 10^{17}/cm^3, find the hole and electron concentrations at 27°C and 125°C. Is the result n- or p-type silicon?

Section 3.3: Current Flow in Semiconductors

3.7 Find the end-to-end resistance of a bar 15-µm long, 3-µm wide, and 1-µm thick, made of the following materials at 25°C:

(a) intrinsic silicon
(b) n-doped silicon with $N_D = 5 \times 10^{16}$/cm^3

problems with blue numbers are considered essential; * = difficult problem; ** = more difficult; *** = very challenging
SIM = simulation; D = design problem; ● = see related video example

Table P3.12

Doping Concentration (carriers/cm^3)	μ_n (cm^2/V · s)	μ_p (cm^2/V · s)	D_n (cm^2/s)	D_p (cm^2/s)
Intrinsic	1350	480		
10^{16}	1200	400		
10^{17}			20	7
10^{18}			10	4

(c) n-doped silicon with $N_D = 5 \times 10^{18}$/cm^3
(d) p-doped silicon with $N_A = 5 \times 10^{16}$/cm^3
(e) aluminum with resistivity of 2.8 μΩ·cm

For intrinsic silicon, use the data in Table 3.1 in the Summary Tables supplement found online at www.oup.com/he/sedra-smith8e. For doped silicon, assume $\mu_n = 3\mu_p = 1200$ cm^2/V · s. (Recall that $R = \rho L/A$.)

3.8 Contrast the electron and hole drift velocities through a 10-μm layer of intrinsic silicon across which a voltage of 3 V is imposed. Let μ_n = 1350 cm^2/V · s and μ_p = 480 cm^2/V · s·

3.9 Find the length of a silicon bar having a 5-μm × 4-μm cross section and having free-electron and hole densities of 10^4/cm^3 and 10^{16}/cm^3, respectively, so that 0.2 mA current flows when 1 V is applied end-to-end. Use μ_n = 1200 cm^2/V · s and μ_p = 500 cm^2/V · s.

3.10 In a 10-μm-long bar of donor-doped silicon, what donor concentration is needed to realize a current density of 2 mA/μm^2 in response to an applied voltage of 1 V? (*Note*: Although the carrier mobilities change with doping concentration, as a first approximation you may assume μ_n to be constant and use 1350 cm^2/V · s, the value for intrinsic silicon.)

3.11 Holes are being steadily injected into a region of n-type silicon (connected to other devices, the details of which are not important for this question). In the steady state, the excess-hole concentration profile shown in Fig. P3.11 is established in the n-type silicon region at room temperature. Here "excess" means over and above the thermal-equilibrium concentration (in the absence of hole injection), denoted p_{n0}. If $N_D = 10^{16}$/cm^3, $n_i = 1.5 \times 10^{10}$/cm^3, D_p = 12 cm^2/s, and W = 50 nm, find the density of the current that will flow in the x direction.

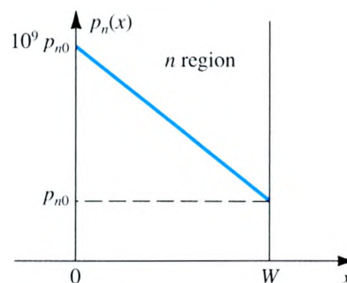


Figure P3.11

3.12 Both the carrier mobility and the diffusivity decrease as the doping concentration of silicon is increased. Table P3.12 provides a few data points for μ_n, μ_p, D_n, and D_p versus doping concentration at room temperature. Use the Einstein relationship to find the missing values.

Section 3.4: The *pn* Junction

3.13 Calculate the built-in voltage of a junction in which the p and n regions are doped equally with 5×10^{16} atoms/cm^3. Assume $n_i = 1.5 \times 10^{10}$/cm^3. With the terminals left open, what is the width of the depletion region, and how far does it extend into the p and n regions? If the cross-sectional area of the junction is 2 μm^2, find the magnitude of the charge stored on either side of the junction.

3.14 If, for a particular junction, the acceptor concentration is 10^{17}/cm^3 and the donor concentration is 10^{16}/cm^3, find the junction built-in voltage. Assume $n_i = 1.5 \times 10^{10}$/cm^3. Also, find the width of the depletion region (W) and its extent in each of the p and n regions when the junction terminals are left open. Calculate the magnitude of the charge stored on either side of the junction. Assume that the junction area is 10 μm^2.

3.15 Estimate the total charge stored in a 0.1-μm depletion layer on one side of a 10-μm × 10-μm junction. The doping concentration on that side of the junction is 5×10^{17}/cm^3.

3.16 In a pn junction for which $N_A \gg N_D$, and the depletion layer exists mostly on the lightly doped side with W = 0.2 μm, find V_0 if $N_D = 10^{16}$/cm^3. Also calculate Q_J for the case A = 10 μm^2.

3.17 By how much does V_0 change if N_A or N_D is increased by a factor of 10?

3.18 In a *pn* junction with $N_A \gg N_D \gg n_i$, sketch a plot of charge density on both sides of the junction. How does the sketch change if N_D is increased by a factor of 4?

3.19 Given that the electric field E has a triangular profile as shown in Fig. P3.19, and that the built-in potential, V_0, is given by the total area under the triangle, find an expression for the maximum electric field magnitude, E_{max}, in terms of the dopant concentrations and fundamental constants.

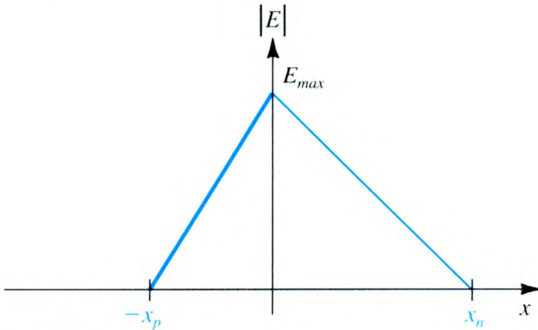


Figure P3.19

Section 3.5: The *pn* Junction with an Applied Voltage

3.20 If a 2-V reverse-bias voltage is applied across the junction specified in Problem 3.14, find W and Q_J.

3.21 Show that for a *pn* junction reverse-biased with a voltage V_R, the depletion-layer width W and the charge stored on either side of the junction, Q_J, can be expressed as

$$W = W_0\sqrt{1+\frac{V_R}{V_0}}$$

$$Q_J = Q_{J0}\sqrt{1+\frac{V_R}{V_0}}$$

where W_0 and Q_{J0} are the values in equilibrium.

3.22 In a forward-biased *pn* junction show that the ratio of the current component due to hole injection across the junction to the component due to electron injection is given by

$$\frac{I_p}{I_n} = \frac{D_p}{D_n}\frac{L_n}{L_p}\frac{N_A}{N_D}$$

Evaluate this ratio for the case $N_A = 10^{18}/\text{cm}^3$, $N_D = 10^{16}/\text{cm}^3$, $L_p = 5\ \mu\text{m}$, $L_n = 10\ \mu\text{m}$, $D_p = 10\ \text{cm}^2/\text{s}$, and $D_n = 20\ \text{cm}^2/\text{s}$, and hence find I_p and I_n for the case in which the *pn* junction is conducting a forward current $I = 100\ \mu\text{A}$.

3.23 Calculate I_S and the current I for $V = 780$ mV for a *pn* junction for which $N_A = 10^{17}/\text{cm}^3$, $N_D = 10^{16}/\text{cm}^3$, $A = 20\ \mu\text{m}^2$, $n_i = 1.5\times10^{10}/\text{cm}^3$, $L_p = 5\ \mu\text{m}$, $L_n = 10\ \mu\text{m}$, $D_p = 10\ \text{cm}^2/\text{s}$, and $D_n = 18\ \text{cm}^2/\text{s}$.

3.24 Assuming that the temperature dependence of I_S arises mostly because I_S is proportional to n_i^2, use the expression for n_i in Eq. (3.2) to determine the factor by which n_i^2 changes as T changes from 300 K to 310 K. This will be approximately the same factor by which I_S changes for a 10°C rise in temperature. What is the factor?

3.25 A p^+n junction is one in which the doping concentration in the *p* region is much greater than that in the *n* region. In such a junction, the forward current is mostly due to hole injection across the junction. Show that

$$I \simeq I_p = Aqn_i^2\frac{D_p}{L_pN_D}\left(e^{V/V_T}-1\right)$$

For the specific case in which $N_D = 10^{16}/\text{cm}^3$, $D_p = 10\ \text{cm}^2/\text{s}$, $L_p = 10\ \mu\text{m}$, and $A = 10^4\ \mu\text{m}^2$, find I_S and the voltage V obtained when $I = 0.2$ mA. Assume operation at 300 K where $n_i = 1.5\times10^{10}/\text{cm}^3$.

3.26 A *pn* junction for which the breakdown voltage is 12 V has a rated (i.e., maximum allowable) power dissipation of 0.25 W. What continuous current in the breakdown region will raise the dissipation to half the rated value? If breakdown occurs for only 10 ms in every 20 ms, what average breakdown current is allowed?

Section 3.6: Capacitive Effects in the *pn* Junction

3.27 For the *pn* junction specified in Problem 3.14, find C_{j0} and C_j at $V_R = 2$ V.

3.28 For a particular junction for which $C_{j0} = 0.4$ pF, $V_0 = 0.75$ V, and $m = 1/3$, find C_j at reverse-bias voltages of 1 V and 3 V.

3.29 The junction capacitance C_j can be thought of as that of a parallel-plate capacitor and thus given by

$$C_j = \frac{\epsilon A}{W}$$

Show that this approach leads to a formula identical to that obtained by combining Eqs. (3.42) and (3.44) [or equivalently, by combining Eqs. (3.44) and (3.45)].

3.30 A *pn* junction operating in the forward-bias region with a current I of 0.4 mA is found to have a diffusion capacitance of 1 pF. What diffusion capacitance do you expect this junction to have at $I = 0.1$ mA? What is the mean transit time for this junction?

***3.31** A **short-base diode** is one where the widths of the p and n regions are much smaller than L_n and L_p, respectively. As a result, the excess minority-carrier distribution in each region is a straight line rather than the exponentials shown in Fig. 3.12.

(a) For the short-base diode, sketch a figure corresponding to Fig. 3.12 and assume as in Fig. 3.12 that $N_A \gg N_D$.

(b) Following a derivation similar to that given in Section 3.5.2, show that if the widths of the p and n regions are denoted W_p and W_n then

$$I = Aqn_i^2\left[\frac{D_p}{(W_n - x_n)N_D} + \frac{D_n}{(W_p - x_p)N_A}\right]\left(e^{V/V_T} - 1\right)$$

and

$$Q_p = \frac{1}{2}\frac{(W_n - x_n)^2}{D_p}I_p$$

$$\simeq \frac{1}{2}\frac{W_n^2}{D_p}I_p, \text{ for } W_n \gg x_n$$

(c) Also, assuming $Q \simeq Q_p$, $I \simeq I_p$, show that

$$C_d = \frac{\tau_T}{V_T}I$$

where

$$\tau_T = \frac{1}{2}\frac{W_n^2}{D_p}$$

(d) If a designer wishes to limit C_d to 8 pF at $I = 1$ mA, what should W_n be? Assume $D_p = 10\ \text{cm}^2/\text{s}$.

CHAPTER 4

Diodes

IN THIS CHAPTER YOU WILL LEARN

- The characteristics of the ideal diode and how to analyze and design circuits containing multiple ideal diodes together with resistors and dc sources to realize useful and interesting nonlinear functions.
- The details of the i–v characteristic of the junction diode and how to use it to analyze diode circuits operating in the various bias regions: forward, reverse, and breakdown.
- A simple but effective model of the diode i–v characteristic in the forward direction: the constant-voltage-drop model.
- A powerful technique for the application and modeling of nonlinear devices such as the diode: dc-biasing the diode and modeling its operation for small signals around the dc operating point by means of the small-signal model.
- Application of the diode in the design of rectifier circuits, which convert ac voltages to dc as needed for powering electronic equipment.
- Other practical and important applications of diodes such as for protecting circuits from excessive voltages and for lighting.

Introduction

In Chapters 1 and 2 we dealt almost entirely with linear circuits; any nonlinearity, such as that introduced by amplifier output saturation, was treated as a problem to be solved by the circuit designer. However, there are many other signal-processing functions that can be implemented only by nonlinear circuits. Examples include the generation of dc voltages from the ac power supply, and the generation of signals of various waveforms (e.g., sinusoids, square waves, pulses). Also, digital logic and memory circuits constitute a special class of nonlinear circuits.

The simplest and most fundamental nonlinear circuit element is the diode. Just like a resistor, the diode has two terminals; but unlike the resistor, which has a linear (straight-line) relationship between the current flowing through it and the voltage appearing across it, the diode has a nonlinear *i–v* characteristic.

In order to understand the essence of the diode function, we begin with a fictitious element, the ideal diode. We then introduce the silicon junction diode, explain its terminal characteristics, and provide techniques for analyzing diode circuits. These techniques rely on device modeling. Our study of modeling the diode characteristics will lay the foundation for modeling transistor operation in the next three chapters.

Diodes are commonly used in the design of rectifiers, which convert ac to dc in many electronic systems. We will thus study rectifier circuits in detail. We will also briefly consider other applications of diodes, including voltage limiting and regulation.

The junction diode is nothing more than the *pn* junction we studied in Chapter 3, and most of this chapter is concerned with the study of silicon *pn*-junction diodes. In the last section, however, we briefly consider some specialized diode types, including the photodiode and the light-emitting diode (LED).

4.1 The Ideal Diode

4.1.1 Current–Voltage Characteristic

The ideal diode can be considered the most fundamental nonlinear circuit element. It is a two-terminal device with the circuit symbol of Fig. 4.1(a) and the *i–v* characteristic shown in Fig. 4.1(b). We can interpret the terminal characteristic of the ideal diode as follows: If a negative voltage (relative to the reference direction indicated in Fig. 4.1a) is applied to the diode, no current flows and the diode behaves as an open circuit (Fig. 4.1c). Diodes operated in this mode are said to be **reverse biased**, or operated in the reverse direction. An ideal diode has zero current when operated in the reverse direction and is said to be **cut off**, or simply **off**.

On the other hand, if a positive current (relative to the reference direction indicated in Fig. 4.1a) is applied to the ideal diode, zero voltage drop appears across the diode. In other words, the ideal diode behaves as a short circuit in the *forward* direction (Fig. 4.1d); it passes any current with zero voltage drop. A **forward-biased** diode is said to be **turned on**, or simply **on**.

From this description we can see that the external circuit must be designed to limit the forward current through a conducting diode, and the reverse voltage across a cutoff diode,

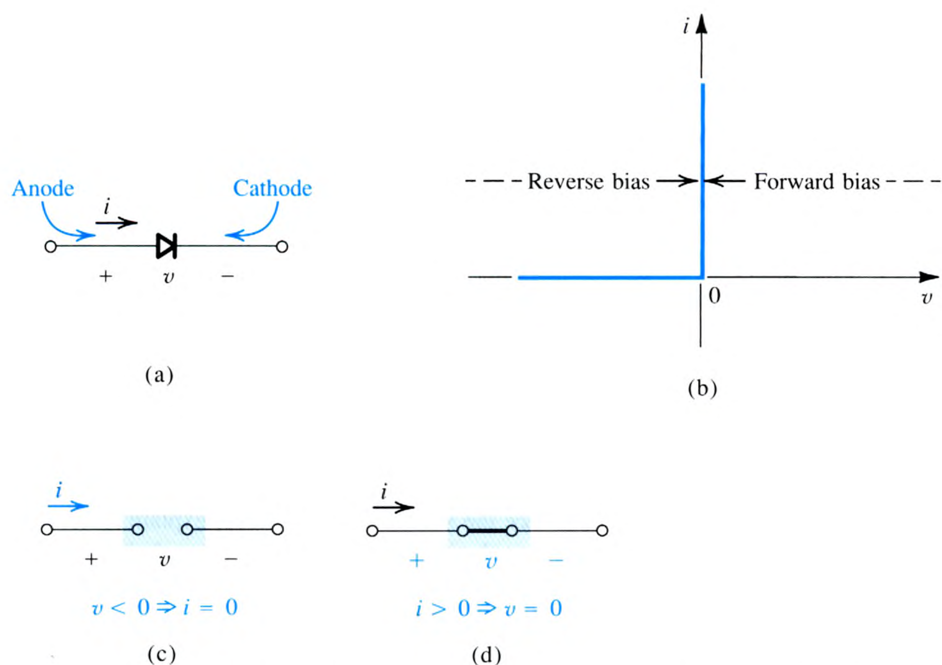


Figure 4.1 The ideal diode: **(a)** diode circuit symbol; **(b)** i–v characteristic; **(c)** equivalent circuit in the reverse direction; **(d)** equivalent circuit in the forward direction.

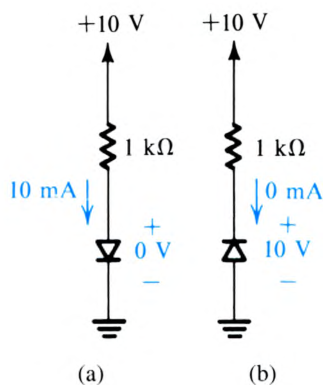


Figure 4.2 The two modes of operation of ideal diodes and the use of an external circuit to limit **(a)** the forward current and **(b)** the reverse voltage.

to predetermined values. Figure 4.2 shows two diode circuits that illustrate this point. In the circuit of Fig. 4.2(a) the diode is obviously conducting. Thus its voltage drop will be zero, and the current through it will be determined by the +10-V supply and the 1-kΩ resistor as 10 mA. The diode in Fig. 4.2(b) is obviously cut off, and thus its current will be zero, which in turn means that the entire 10-V supply will appear as reverse bias across the diode.

The positive terminal of the diode is called the **anode** and the negative terminal the **cathode**, a carryover from the days of vacuum-tube diodes. The i–v characteristic of the ideal diode (conducting in one direction and not in the other) should explain the choice of its arrow-like circuit symbol.

As you can tell from the preceding description, the i–v characteristic of the ideal diode is highly nonlinear; although it consists of two straight-line segments, they are at 90° to one another. A nonlinear curve that consists of straight-line segments is said to be **piecewise linear**. If a device with a piecewise-linear characteristic is used in a particular application in such a way that the signal across its terminals swings along only one of the linear segments, then the device can be considered a linear circuit element as far as that particular circuit application is concerned. On the other hand, if signals swing past one or more of the break points in the characteristic, linear analysis is no longer possible.

4.1.2 The Rectifier

A fundamental application of the diode, one that makes use of its severely nonlinear i–v curve, is the rectifier circuit shown in Fig. 4.3(a). The circuit consists of the series

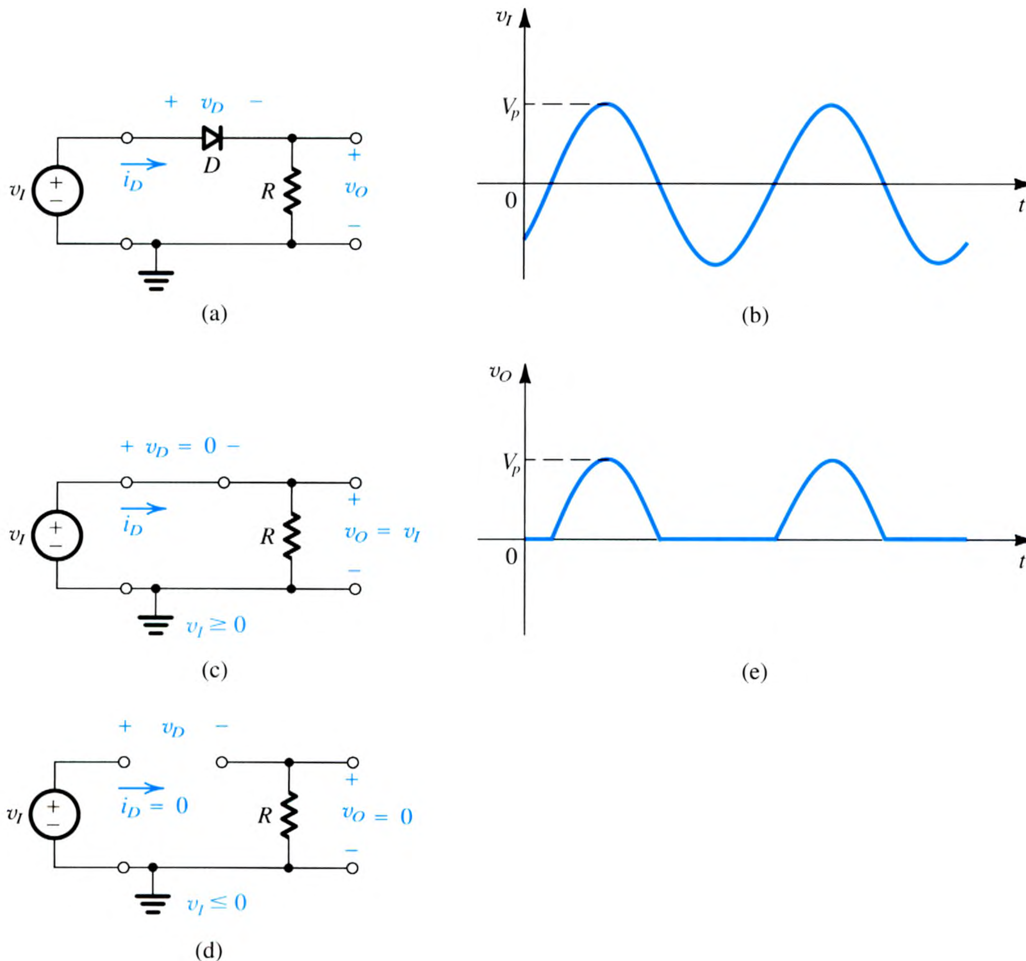


Figure 4.3 **(a)** Rectifier circuit. **(b)** Input waveform. **(c)** Equivalent circuit when $v_I \geq 0$. **(d)** Equivalent circuit when $v_I \leq 0$. **(e)** Output waveform.

connection of a diode D and a resistor R. Let the input voltage v_I be the sinusoid shown in Fig. 4.3(b), and assume the diode to be ideal. During the positive half-cycles of the input sinusoid, the positive v_I will cause current to flow through the diode in its forward direction. It follows that the diode voltage v_D will be very small—ideally zero. Thus the circuit will have the equivalent shown in Fig. 4.3(c), and the output voltage v_O will be equal to the input voltage v_I. On the other hand, during the negative half-cycles of v_I, the diode will not conduct. Thus the circuit will have the equivalent shown in Fig. 4.3(d), and v_O will be zero. Thus the output voltage will have the waveform shown in Fig. 4.3(e). Note that while v_I alternates in polarity and has a zero average value, v_O is unidirectional and has a finite average value or a *dc component*. Thus the circuit of Fig. 4.3(a) **rectifies** the signal and hence is called a **rectifier**. It can be used to generate dc from ac. We will study rectifier circuits in Section 4.6.

EXERCISES

4.1 For the circuit in Fig. 4.3(a), sketch the transfer characteristic v_O versus v_I.
Ans. See Fig. E4.1

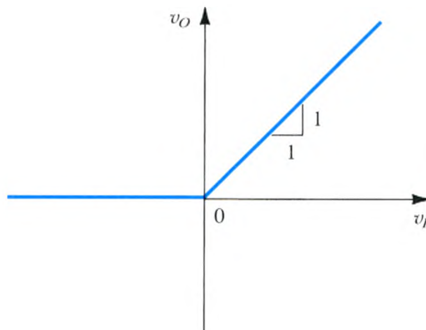


Figure E4.1

4.2 For the circuit in Fig. 4.3(a), sketch the waveform of v_D.
Ans. $v_D = v_I - v_O$, resulting in the waveform in Fig. E4.2

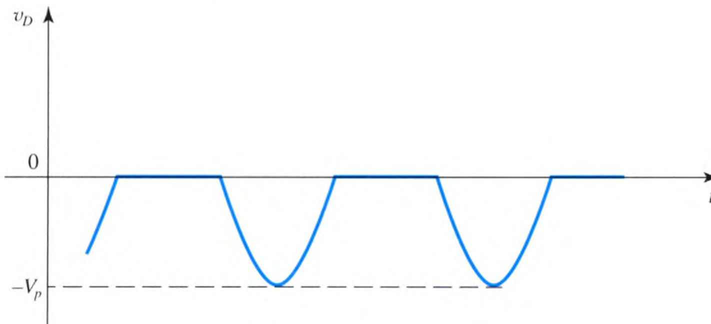


Figure E4.2

4.3 In the circuit of Fig. 4.3(a), let v_I have a peak value of 10 V and $R = 5$ kΩ. Find the peak value of i_D and the dc component of v_O. (*Hint*: The average value of half-sine waves is V_p/π.)
Ans. 2 mA; 3.18 V

Example 4.1

Figure 4.4(a) shows a circuit for charging a 12-V battery. If v_S is a sinusoid with 24-V peak amplitude, find the fraction of each cycle during which the diode conducts. Also, find the peak value of the diode current and the maximum reverse-bias voltage that appears across the diode.

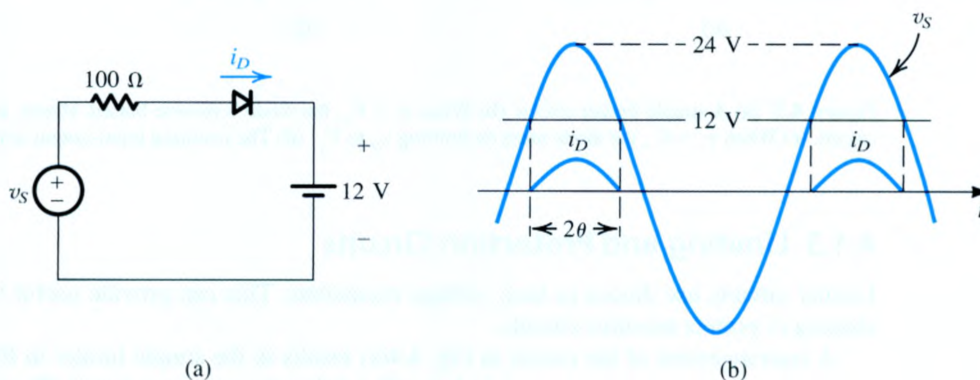


Figure 4.4 Circuit and waveforms for Example 4.1.

Solution

The diode conducts when v_S exceeds 12 V, as shown in Fig. 4.4(b). The conduction angle is 2θ, where θ is given by

$$24\cos\theta = 12$$

Thus $\theta = 60°$ and the conduction angle is 120°, or one-third of a cycle.

The peak value of the diode current is given by

$$I_d = \frac{24 - 12}{100} = 0.12 \text{ A}$$

The maximum reverse voltage across the diode occurs when v_S is at its negative peak and is equal to $24 + 12 = 36$ V.

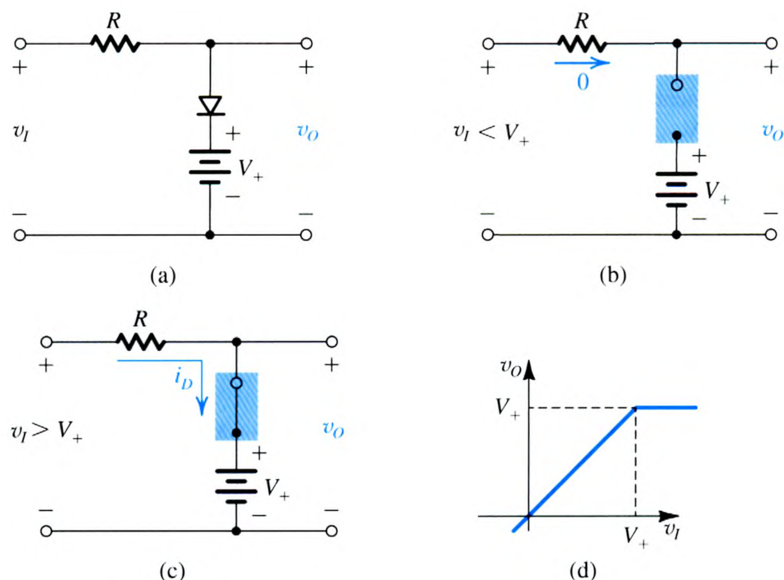


Figure 4.5 (**a**) A simple limiter circuit. (**b**) When $v_I < V_+$, the diode is reverse biased. Hence, it is an open circuit. (**c**) When $v_I > V_+$, the diode turns on limiting v_O to V_+. (**d**) The resulting input-output relationship.

4.1.3 Limiting and Protection Circuits

Limiter circuits use diodes to limit voltage excursions. This can provide useful waveform shaping or protect sensitive circuits.

A rearrangement of the circuit in Fig. 4.4(a) results in the simple limiter in Fig. 4.5(a). When the input v_I is below V_+, the diode is off and therefore an open circuit. The equivalent schematic is shown in Fig. 4.5(b). No current flows through the resistor R, and $v_O = v_I$. When the input v_I exceeds V_+, the diode turns on, resulting in the equivalent circuit of Fig. 4.5(c), where $v_O = V_+$. Thus, the circuit keeps v_O from exceeding the upper limit established by V_+. Its input–output characteristic is shown in Fig. 4.5(d). When on, the diode must conduct a current,

$$i_D = \frac{v_I - V_+}{R}$$

Reversing the diode's polarity results in a limiter that establishes a lower limit on the voltage excursions.

Example 4.2

Assuming ideal diodes, find the input–output characteristic of the limiter circuit in Fig. 4.6(a). If $v_I =$ 100 V and $R = 100\ \Omega$, what current is conducted by each of the two diodes?

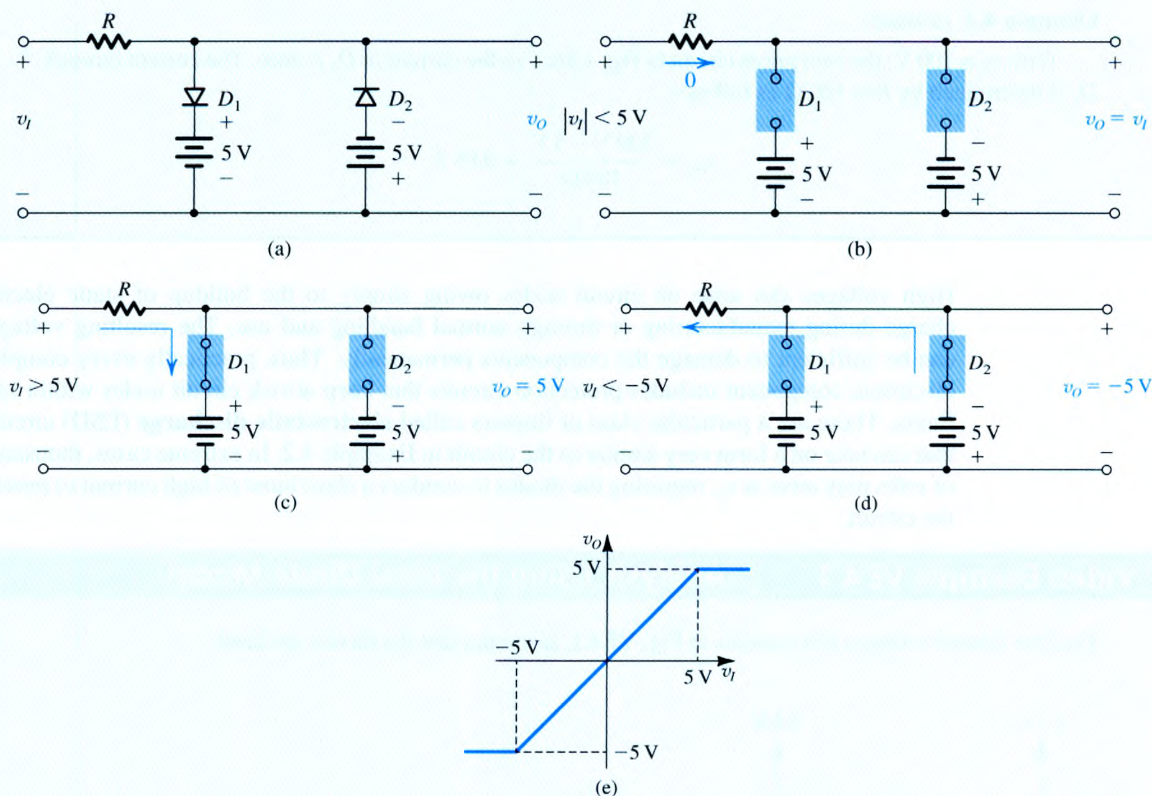


Figure 4.6 (**a**) Simple limiter circuit in Example 4.2. (**b**)–(**d**) The equivalent circuit for differential input voltages. (**e**) The input–output relationship.

Solution

The diodes are both reverse biased when $-5\text{ V} < v_I < +5\text{ V}$. The equivalent circuit in this condition appears in Fig. 4.6(b), where we replace diodes D_1 and D_2 by open circuits, and $v_O = v_I$. When v_I increases beyond +5 V, D_1 turns on while D_2 remains off. We illustrate this situation in Fig. 4.6(c), where v_O is shorted to +5 V and a current $(v_I - 5\text{ V})/R$ flows through the resistor. Finally, when $v_I < -5$ V, D_2 is on while D_1 is off, as shown in Fig. 4.6(d). Here, $v_O = -5$ V while current flows through D_2 and R, into v_I.

In summary, the output voltage of the limiter circuit is given by

$$v_O = \begin{cases} -5\text{ V} , & v_I < -5\text{ V} \\ v_I , & -5\text{ V} < v_I < +5\text{ V} \\ +5\text{ V} , & v_I > +5\text{ V} \end{cases}$$

A plot of v_O versus v_I is presented in Fig. 4.6(e).

Example 4.2 *continued*

With $v_I = 100$ V, the equivalent circuit is Fig. 4.6(c), so the current in D_2 is zero. The current through D_1 is determined by $R = 100\ \Omega$ as follows:

$$i_{D_1} = \frac{100\text{ V} - 5\text{ V}}{100\ \Omega} = 0.95\text{ A}$$

High voltages can arise on circuit nodes owing simply to the buildup of static electric charge during manufacturing or through normal handling and use. The resulting voltages can be sufficient to damage the components permanently. Thus, practically every complex electronic component includes protection circuits that keep at-risk circuit nodes within safe limits. These are a particular class of limiters called **electrostatic discharge** (ESD) circuits that can take on a form very similar to the circuit in Example 4.2. In extreme cases, thousands of volts may arise at v_I, requiring the diodes to conduct a short burst of high current to protect the circuit.

Video Example VE 4.1 Analysis Using the Ideal Diode Model

Find the labeled voltages and currents in Fig. VE4.1, assuming that the diodes are ideal.

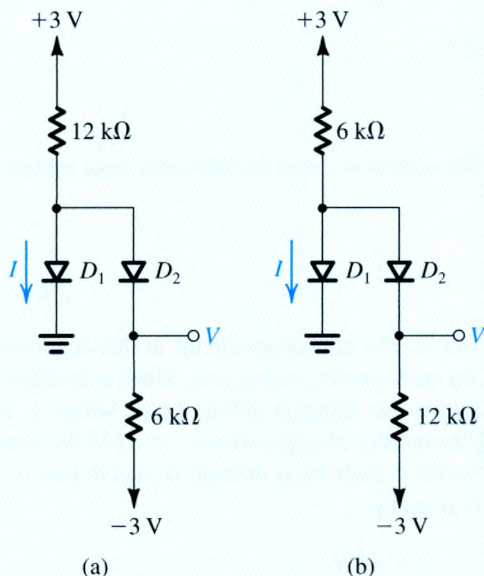


Figure VE4.1 Circuit for Video Example 4.1.


VE 4.1

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 4.7

EXERCISES

4.4 Find the values of I and V in the circuits shown in Fig. E4.4.

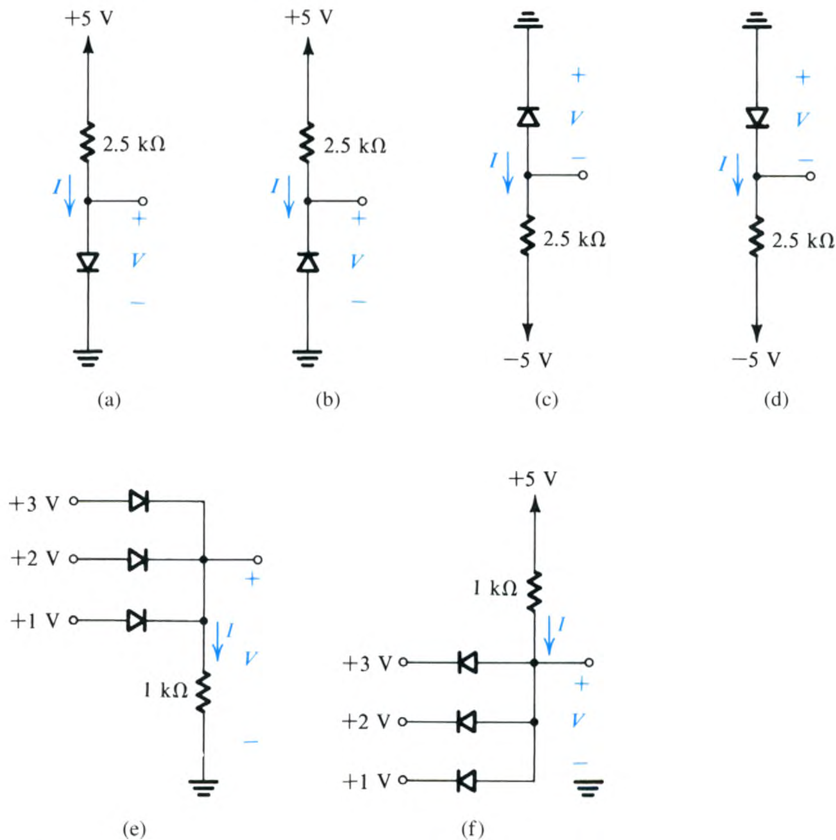


Figure E4.4

Ans. (a) 2 mA, 0 V; (b) 0 mA, 5 V; (c) 0 mA, 5 V; (d) 2 mA, 0 V; (e) 3 mA, +3 V; (f) 4 mA, +1 V

4.5 Figure E4.5 shows a circuit for an ac voltmeter. It uses a moving-coil meter that gives a full-scale reading when the *average* current flowing through it is 1 mA. The moving-coil meter has a 50-Ω resistance.

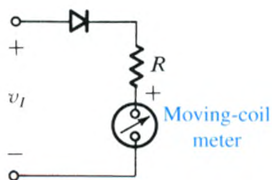


Figure E4.5

Find the value of R that results in the meter indicating a full-scale reading when the input sine-wave voltage v_I is 20 V peak-to-peak. (*Hint:* The average value of half-sine waves is V_p/π.)

Ans. 3.133 kΩ

4.6 Select the resistor value R in Fig. 4.6 so that the maximum diode current is 50 mA for $|v_I| < 20$ V.

Ans. $R = 300\ \Omega$

4.2 Terminal Characteristics of Junction Diodes

The most common implementation of the diode uses a *pn* junction. We studied the physics of the *pn* junction and derived its i–v characteristic in Chapter 3. That the *pn* junction is used to implement the diode function should come as no surprise: the *pn* junction can conduct substantial current in the forward direction and almost no current in the reverse direction. In this section we study the i–v characteristic of the *pn* junction diode in detail in order to prepare ourselves for diode circuit applications.

Figure 4.7 shows the i–v characteristic of a silicon junction diode. The same characteristic is shown in Fig. 4.8 with some scales expanded and others compressed to reveal details. Note that the scale changes have resulted in the apparent discontinuity at the origin.

As indicated, the characteristic curve consists of three distinct regions:

1. The forward-bias region, determined by $v > 0$
2. The reverse-bias region, determined by $v < 0$
3. The breakdown region, determined by $v < -V_{BR}$

These three regions of operation are described in the following sections.

4.2.1 The Forward-Bias Region

The diode enters the forward-bias—or, simply, forward—region of operation when the terminal voltage v is positive. In the forward region the i–v relationship is closely approximated by

$$i = I_S\left(e^{v/V_T} - 1\right) \tag{4.1}$$

In this equation[1] I_S is a constant for a given diode at a given temperature. A formula for I_S in terms of the diode's physical parameters and temperature was given in Eq. (3.40). The current I_S is usually called the **saturation current** (for reasons that will become apparent shortly). Another name for I_S, and one that we will occasionally use, is the **scale current**.

[1]Equation (4.1), the diode equation, is sometimes written to include a constant n in the exponential,

$$i = I_S(e^{v/nV_T} - 1)$$

with n having a value between 1 and 2, depending on the material and the physical structure of the diode. Diodes using the standard integrated-circuit fabrication process exhibit $n = 1$ when operated under normal conditions. For simplicity, we shall use $n = 1$ throughout this book, unless otherwise specified.

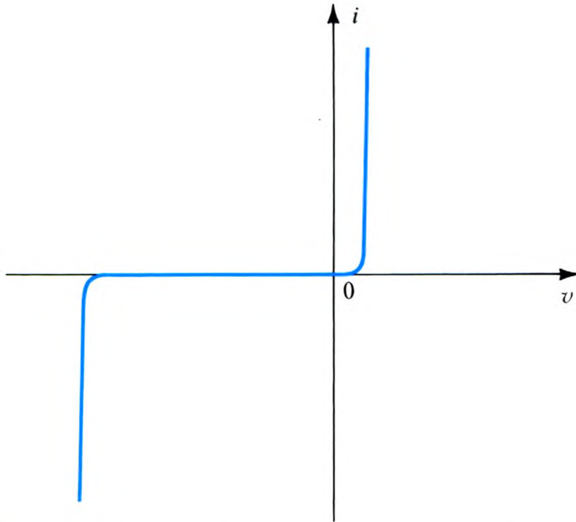


Figure 4.7 The i–v characteristic of a silicon junction diode.

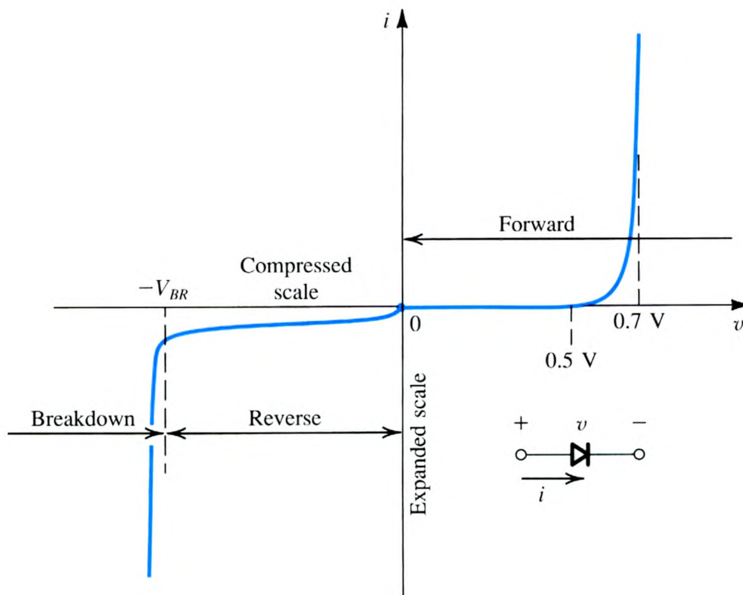


Figure 4.8 The silicon diode i–v relationship with some scales expanded and others compressed in order to reveal details.

This name comes from the fact that I_S is directly proportional to the cross-sectional area of the diode. Doubling the junction area produces a diode with double the value of I_S and, as the diode equation indicates, double the value of current i for a given forward voltage v. For "small-signal" diodes, which are small-size diodes intended for low-power applications, I_S is

on the order of 10^{-15} A. The value of I_S is, however, a very strong function of temperature. As a rule of thumb, I_S doubles in value for every 5°C rise in temperature.

The voltage V_T in Eq. (4.1) is a constant called the **thermal voltage** and is given by

$$V_T = \frac{kT}{q} \tag{4.2}$$

where

k = Boltzmann's constant = 8.62×10^{-5} eV/K = 1.38×10^{-23} joules/kelvin
T = the absolute temperature in kelvins = 273 + temperature in °C
q = the magnitude of electronic charge = 1.60×10^{-19} coulomb

Substituting these into Eq. (4.2) gives

$$V_T = 0.0862T, \text{ mV} \tag{4.2a}$$

Thus, at room temperature (20°C) the value of V_T is 25.3 mV. In rapid approximate circuit analysis we will use $V_T \simeq 25$ mV at room temperature.[2]

For appreciable current i in the forward direction, specifically for $i \gg I_S$, we can approximate Eq. (4.1) by the exponential relationship

$$i \simeq I_S e^{v/V_T} \tag{4.3}$$

Alternatively, we can express this relationship in the logarithmic form:

$$v = V_T \ln \frac{i}{I_S} \tag{4.4}$$

where ln denotes the natural (base e) logarithm.

The exponential relationship of the current i to the voltage v holds over a span of many decades of current (e.g., a factor of 10^7). This is quite a remarkable property of junction diodes, one that is also found in bipolar junction transistors and that has been exploited in many interesting applications.

Let's consider the forward i–v relationship in Eq. (4.3) and evaluate the current I_1 corresponding to a diode voltage V_1:

$$I_1 = I_S e^{V_1/V_T}$$

Similarly, if the voltage is V_2, the diode current I_2 will be

$$I_2 = I_S e^{V_2/V_T}$$

[2]A slightly higher ambient temperature (25°C or so) is usually assumed for electronic equipment operating inside a cabinet. At this temperature, $V_T \simeq 25.8$ mV. Nevertheless, for the sake of simplicity and to promote rapid circuit analysis, we'll use the more arithmetically convenient value of $V_T \simeq 25$ mV throughout this book.

These two equations can be combined to produce

$$\frac{I_2}{I_1} = e^{(V_2 - V_1)/V_T}$$

which can be rewritten as

$$V_2 - V_1 = V_T \ln \frac{I_2}{I_1}$$

or, in terms of base-10 logarithms,

$$V_2 - V_1 = 2.3\, V_T \log \frac{I_2}{I_1} \tag{4.5}$$

This equation states that for a decade (factor of 10) change in current, the diode voltage drop changes by $2.3V_T$, which is roughly 60 mV. This also suggests that the diode i–v relationship is most conveniently plotted on semilog paper. Using the vertical, linear axis for v and the horizontal, log axis for i, we get a straight line with a slope of 60 mV per decade of current.

A glance at the i–v characteristic in the forward region (Fig. 4.8) reveals that the current is negligibly small for v less than about 0.5 V. This value is usually referred to as the **cut-in voltage**. However, this apparent threshold in the characteristic is simply a consequence of the exponential relationship. Another consequence of this relationship is the rapid increase of i. Thus, for a "fully conducting" diode, the voltage drop lies in a narrow range, approximately 0.6 V to 0.8 V. This gives rise to a simple "model" for the diode where we assume that a conducting diode has approximately a 0.7-V drop across it. Diodes with different current ratings (i.e., different areas and correspondingly different I_S) will exhibit the 0.7-V drop at different currents. For instance, a small-signal diode can be considered to have a 0.7-V drop at $i = 1$ mA, while a higher-power diode can have a 0.7-V drop at $i = 1$ A. We will study diode-circuit analysis and diode models in the next section.

Example 4.3

A silicon diode said to be a 1-mA device displays a forward voltage of 0.7 V at a current of 1 mA. Evaluate the junction scaling constant I_S. What scaling constants would apply for a 1-A diode of the same manufacture that conducts 1 A at 0.7 V?

Solution

Since

$$i = I_S e^{v/V_T}$$

then

$$I_S = i e^{-v/V_T}$$

For the 1-mA diode:

$$I_S = 10^{-3} e^{-700/25} = 6.9 \times 10^{-16} \text{ A}$$

Example 4.3 *continued*

The diode conducting 1 A at 0.7 V corresponds to one thousand 1-mA diodes in parallel with a total junction area 1000 times greater. Thus I_S is also 1000 times greater,

$$I_S = 6.9 \times 10^{-13} \text{ A}$$

Video Example VE 4.2 Analysis Using the Exponential Diode Model

In the circuit shown in Fig. VE4.2, D_1 has 10 times the junction area of D_2. What value of V results? To obtain a value for V of 60 mV, what current I_2 is needed?

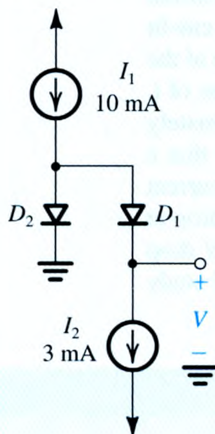


Figure VE4.2 Circuit for Video Example 4.2.



VE 4.2

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 4.27

Diodes made from different semiconductor materials may exhibit values of I_S differing by several orders of magnitude. As a result, for example, germanium diodes can conduct appreciable current at much lower voltages, around 0.3 V. Light-emitting diodes appear to turn on at a much higher voltage, typically over 1 V depending on the color of light. Unless otherwise stated, we focus on silicon diodes that have an approximately 0.7-V drop when conducting.

Since for any given diode both I_S and V_T are functions of temperature, the forward i–v characteristic varies with temperature, as illustrated in Fig. 4.9. At a given constant diode current, the voltage drop across the diode decreases by about 2 mV for every 1°C increase in

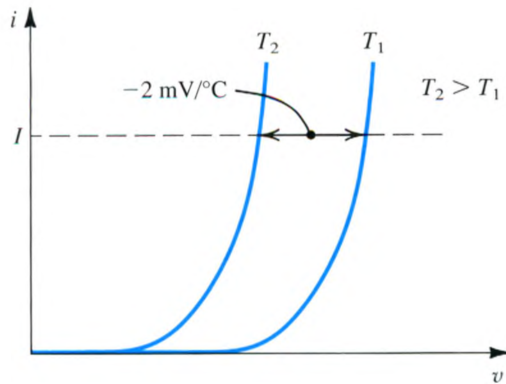


Figure 4.9 Temperature dependence of the diode forward characteristic. At a constant current, the voltage drop decreases by approximately 2 mV for every 1°C increase in temperature.

temperature. The change in diode voltage with temperature has been exploited in the design of electronic thermometers.

EXERCISES

4.7 Find the change in diode voltage if the current changes from 0.1 mA to 10 mA.
Ans. 120 mV

4.8 A silicon junction diode has $v = 0.7$ V at $i = 1$ mA. Find the voltage drop at $i = 0.1$ mA and $i = 10$ mA.
Ans. 0.64 V; 0.76 V

4.9 A germanium diode exhibits a forward drop of 0.3 V while conducting 0.2 mA. Find its scale current, I_S.
Ans. $I_S = 1.23 \times 10^{-9}$ A

4.10 Using the fact that a silicon diode has $I_S = 10^{-14}$ A at 25°C and that I_S increases by 15% per °C rise in temperature, find the value of I_S at 125°C.
Ans. 1.17×10^{-8} A

4.2.2 The Reverse-Bias Region

The diode enters the reverse-bias region of operation when the voltage v is made negative. Equation (4.1) predicts that if v is negative and a few times larger than V_T (25 mV) in magnitude, the exponential term becomes negligibly small compared to unity, and the diode current becomes

$$i \simeq -I_S$$

That is, the current in the reverse direction is constant and equal to I_S. This is where the term *saturation current* comes from.

Real diodes exhibit reverse currents that, though quite small, are much larger than I_S. For instance, a small-signal diode whose I_S is roughly 10^{-14} A to 10^{-15} A could show a reverse current on the order of 1 nA. The reverse current also increases somewhat with the increase in magnitude of the reverse voltage. Note that because the current is so small, these details are not clearly evident on the diode *i–v* characteristic of Fig. 4.8.

A large part of the reverse current is due to leakage effects. These leakage currents are proportional to the junction area, just as I_S is. Their dependence on temperature, however, is different from that of I_S. Thus, whereas I_S doubles for every 5°C rise in temperature, the corresponding rule of thumb for the temperature dependence of the reverse current is that it doubles for every 10°C rise in temperature.

EXERCISE

4.11 The diode in the circuit of Fig. E4.11 is a large high-current device whose reverse leakage is reasonably independent of voltage. If $V = 1$ V at 20°C, find the value of V at 40°C and at 0°C.

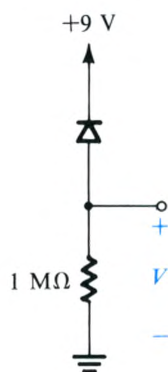


Figure E4.11

Ans. 4 V; 0.25 V

4.2.3 The Breakdown Region

The third distinct region of diode operation is the breakdown region, which is easy to identify on the diode *i–v* characteristic in Fig. 4.8. The diode enters the breakdown region when the magnitude of the reverse voltage exceeds a threshold value that is specific to the particular diode, called the **breakdown voltage** and denoted V_{BR}.

As we can see from Fig. 4.8, in the breakdown region the reverse current increases rapidly, and the increase in voltage drop is very small. Diode breakdown is normally not destructive, provided the power dissipated in the diode is limited by external circuitry to a "safe" level. This safe value is normally specified on the device data sheets. It therefore is

necessary to limit the reverse current in the breakdown region to a value consistent with the permissible power dissipation. Some useful applications of reverse breakdown include voltage regulation (Section 4.5) and light detection (Section 4.7.4).

4.3 Modeling the Diode

Having studied the diode terminal characteristics, we can begin analyzing circuits using forward-conducting diodes, such as the one in Fig. 4.10. It consists of a dc source V_{DD}, a resistor R, and a diode. To determine the diode voltage V_D and current I_D, there are a variety of diode models that could help with our analysis. So far, we know two: the ideal-diode model and the exponential model. We'll assess the suitability of these two models in various analysis situations. We will also develop and comment on other models. This material, besides being useful in the analysis and design of diode circuits, establishes a foundation for modeling transistor operation, which we will study in the next three chapters.

4.3.1 The Exponential Model

The exponential model provides the most accurate description of diode operation outside of the breakdown region. Unfortunately, its severely nonlinear nature makes it the most difficult to use. To illustrate, let's analyze the circuit in Fig. 4.10 using the exponential diode model.

Assuming that V_{DD} is greater than 0.5 V or so, the diode current will be much greater than I_S, and we can represent the diode i–v characteristic by the exponential relationship, resulting in

$$I_D = I_S e^{V_D/V_T} \tag{4.6}$$

The other equation that governs circuit operation is found by writing a Kirchhoff loop equation, giving

$$I_D = \frac{V_{DD} - V_D}{R} \tag{4.7}$$

Assuming that the diode parameter I_S is known, Eqs. (4.6) and (4.7) are two equations in the two unknown quantities I_D and V_D. We can solve for these using either graphical analysis or iterative analysis.

4.3.2 Graphical Analysis Using the Exponential Model

Graphical analysis is performed by plotting the relationships of Eqs. (4.6) and (4.7) on the i–v plane. The solution is the point of intersection of the two graphs. A sketch of the graphical construction is shown in Fig. 4.11. The curve represents the exponential diode equation (Eq. 4.6), and the straight line represents Eq. (4.7). Such a straight line is known as the **load line**, a name that will become more meaningful in later chapters. The load line intersects the

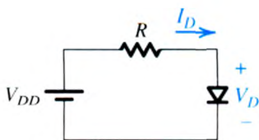


Figure 4.10 A simple circuit used to illustrate the analysis of diode circuits.

diode curve at point Q, which represents the **operating point** of the circuit. Its coordinates give the values of I_D and V_D.

Graphical analysis helps us visualize circuit operation. However, the effort involved in performing such an analysis, particularly for complex circuits, is too great to be justified in practice.

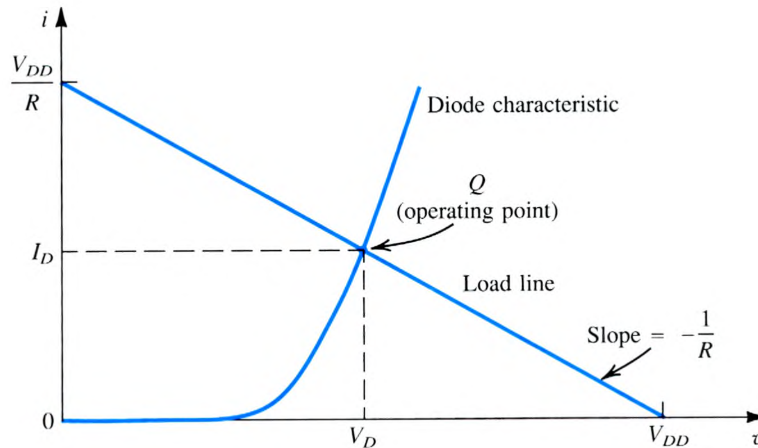


Figure 4.11 Graphical analysis of the circuit in Fig. 4.10 using the exponential diode model.

4.3.3 Iterative Analysis Using the Exponential Model

Equations (4.6) and (4.7) can be solved using a simple iterative procedure illustrated in the following example.

Example 4.4

Determine the current I_D and the diode voltage V_D for the circuit in Fig. 4.10 with $V_{DD} = 5$ V and $R = 1$ kΩ. Assume that the diode has a current of 1 mA at a voltage of 0.7 V.

Solution

To begin the iteration, we assume that $V_D = 0.7$ V and use Eq. (4.7) to determine the current,

$$I_D = \frac{V_{DD} - V_D}{R} = \frac{5 - 0.7}{1} = 4.3 \text{ mA}$$

We then use the diode equation to get a better estimate for V_D. We can do this by using Eq. (4.5),

$$V_2 - V_1 = 2.3 V_T \log \frac{I_2}{I_1}$$

Substituting $2.3V_T = 60$ mV, we have

$$V_2 = V_1 + 0.06 \log \frac{I_2}{I_1}$$

Substituting $V_1 = 0.7$ V, $I_1 = 1$ mA, and $I_2 = 4.3$ mA results in $V_2 = 0.738$ V. Thus the results of the first iteration are $I_D = 4.3$ mA and $V_D = 0.738$ V. The second iteration proceeds in a similar manner:

$$I_D = \frac{5 - 0.738}{1} = 4.262 \text{ mA}$$

$$V_2 = 0.738 + 0.06 \log \left[\frac{4.262}{4.3} \right]$$

$$= 0.738 \text{ V}$$

Thus the second iteration yields $I_D = 4.262$ mA and $V_D = 0.738$ V. Since these values are very close to the values found in the first iteration, no further iterations are necessary, and the solution is $I_D = 4.262$ mA and $V_D = 0.738$ V.

4.3.4 The Need for Rapid Analysis

The iterative analysis procedure used in the example above is simple and yields accurate results after two or three iterations. Nevertheless, there are situations in which the effort and time required are still too much. If we're doing a pencil-and-paper design of a relatively complex circuit, we need rapid and simple circuit analysis. With quick analysis, we can evaluate various possibilities before deciding on a suitable circuit design. To speed up the analysis, we have to accept less precise results. This is seldom a problem, because we can do the more accurate analysis once we have a final or almost-final design with the aid of a computer circuit-analysis program such as SPICE (see Appendix B and the website). We can then use the results of such an analysis to further refine or "fine-tune" the design.

To speed up the analysis, we must find a simpler model for the diode forward characteristic.

4.3.5 The Constant-Voltage-Drop Model

The simplest and most widely used diode model is the constant-voltage-drop model. This model is based on the observation that a forward-conducting diode has a voltage drop that varies in a relatively narrow range, say, 0.6 to 0.8 V. It assumes this voltage to be constant at 0.7 V.[3] With a forward voltage below 0.7 V, it assumes zero diode current. The resulting piecewise-linear model is illustrated in Fig. 4.12.

[3]Different forward voltages may be assumed depending on the diode type. For example, a high-power diode may conduct high currents, giving rise to higher voltages. Diodes made from semiconductor materials other than silicon may exhibit different voltage drops.

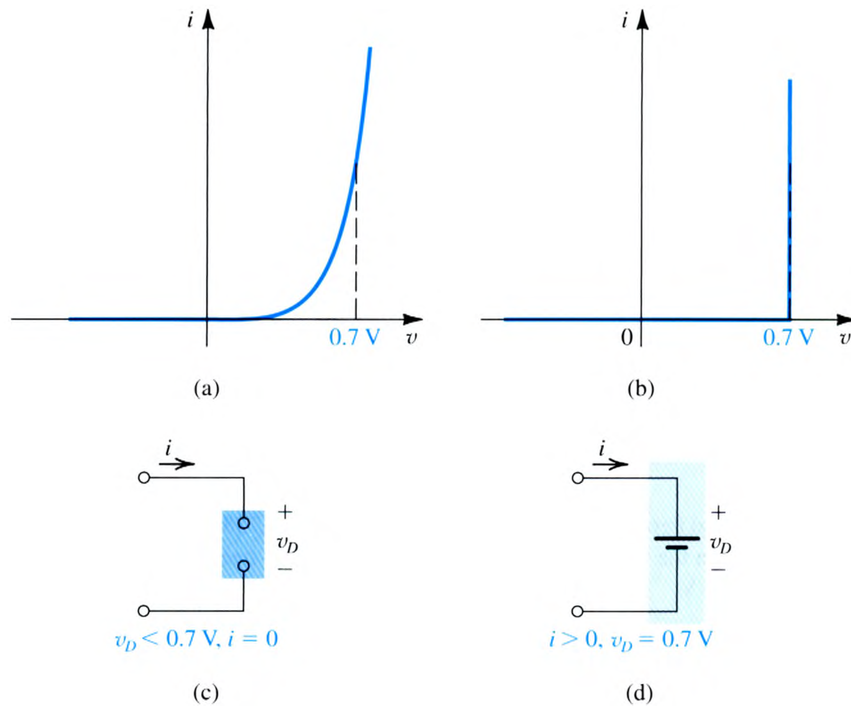


Figure 4.12 Development of the diode constant-voltage-drop model: **(a)** the exponential characteristic; **(b)** approximating the exponential characteristic by a piecewise-linear one; **(c)** the resulting model of the diode with reverse bias or $v_D < 0.7$ V; (d) model of the forward-conducting diode.

The constant-voltage-drop model is the one we use most often in the initial phases of analysis and design. This is especially true if at these stages we do not have detailed information about the diode characteristics, which is often the case.

Finally, note that if we use the constant-voltage-drop model to solve the problem in Example 4.4, we get

$$V_D = 0.7 \text{ V}$$

and

$$I_D = \frac{V_{DD} - 0.7}{R}$$

$$= \frac{5 - 0.7}{1} = 4.3 \text{ mA}$$

which are not very different from the values we got with the more elaborate exponential model.

Video Example VE 4.3 Analysis Using the Constant-Voltage-Drop Diode Model

For the circuits in Fig. VE4.1, using the constant-voltage-drop ($V_D = 0.7$ V) diode model, find the values of the labeled currents and voltages.

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

VE 4.3

Related end-of-chapter problem: 4.43

4.3.6 The Ideal-Diode Model

In applications that involve voltages much greater than the diode voltage drop (0.6 V–0.8 V), we may neglect the diode voltage drop altogether while calculating the diode current. The result is the ideal-diode model, which we studied in Section 4.1. For the circuit in Example 4.4 (i.e., Fig. 4.10 with $V_{DD} = 5$ V and $R = 1$ kΩ), using the ideal-diode model leads to

$$V_D = 0 \text{ V}$$

$$I_D = \frac{5-0}{1} = 5 \text{ mA}$$

which for a very quick analysis would not be bad as a gross estimate. However, with almost no additional work, the 0.7-V-drop model gives us much more realistic results. We note, however, that the greatest utility of the ideal-diode model is in determining which diodes are on and which are off in a multidiode circuit, such as those considered in Section 4.6.

EXERCISES

4.12 For the circuit in Fig. 4.10, find I_D and V_D for the case $V_{DD} = 5$ V and $R = 10$ kΩ. Assume that the diode has a voltage of 0.7 V at 1-mA current. Use (a) iteration and (b) the constant-voltage-drop model with $V_D = 0.7$ V.
Ans. (a) 0.43 mA, 0.68 V; (b) 0.43 mA, 0.7 V

D4.13 Design the circuit in Fig. E4.13 to provide an output voltage of 2.4 V. Assume that the diodes available have 0.7-V drop at 1 mA.
Ans. $R = 139$ Ω

4.14 Repeat Exercise 4.4 using the 0.7-V-drop model to obtain better estimates of I and V than those found in Exercise 4.4 (using the ideal-diode model).
Ans. (a) 1.72 mA, 0.7 V; (b) 0 mA, 5 V; (c) 0 mA, 5 V; (d) 1.72 mA, 0.7 V; (e) 2.3 mA, +2.3 V; (f) 3.3 mA, +1.7 V

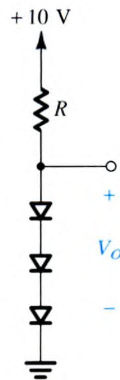


Figure E4.13

4.3.7 Operation in the Reverse Breakdown Region

The very steep i–v curve that the diode exhibits in the breakdown region (Fig. 4.8) and the almost-constant voltage drop that this indicates suggest that we can use diodes operating in the breakdown region for amplification and in the design of circuits whose purpose is to maintain nearly constant dc voltages, called voltage regulators (studied in Section 4.5). The two reverse breakdown mechanisms were introduced in Chapter 3: avalanche and zener. The diode's construction will determine which of these two mechanisms is predominant.

Avalanche Breakdown Avalanche is typically observed with $V_{BR} \gtrsim 10$ V. It is the more common form of reverse breakdown for diodes required to withstand a high reverse voltage without appreciable current. Once operating in avalanche breakdown, the combination of large reverse current at high V_{BR} results in very high power dissipation. Unless we limit the reverse current, overheating can result in permanent damage to the diode. Thus, in most applications we avoid avalanche breakdown by respecting the maximum reverse voltage[4] specified by the diode manufacturer.

In some applications, however, we can take advantage of avalanche operation to translate very small stimuli into large currents. For example, avalanche photodiodes (APDs) are designed to operate in breakdown so that even very low levels of light initiate detectable reverse current. In some APDs the absorption of a single photon sets off a spike of current. These can be used to detect the precise arrival time of weak light pulses.

Zener Breakdown and the Zener Diode Manufactured specifically to operate under zener breakdown, **zener diodes** maintain a nearly constant reverse voltage drop, V_Z, over a wide range of reverse currents. Figure 4.13(a) shows the circuit symbol of the zener diode indicating the polarity of V_Z and current flow in normal applications. Figure 4.13(b)

[4]Some data sheets specify the diode's peak inverse voltage (PIV) in reference to its common application in rectifier circuits, which we discuss in Section 4.6.

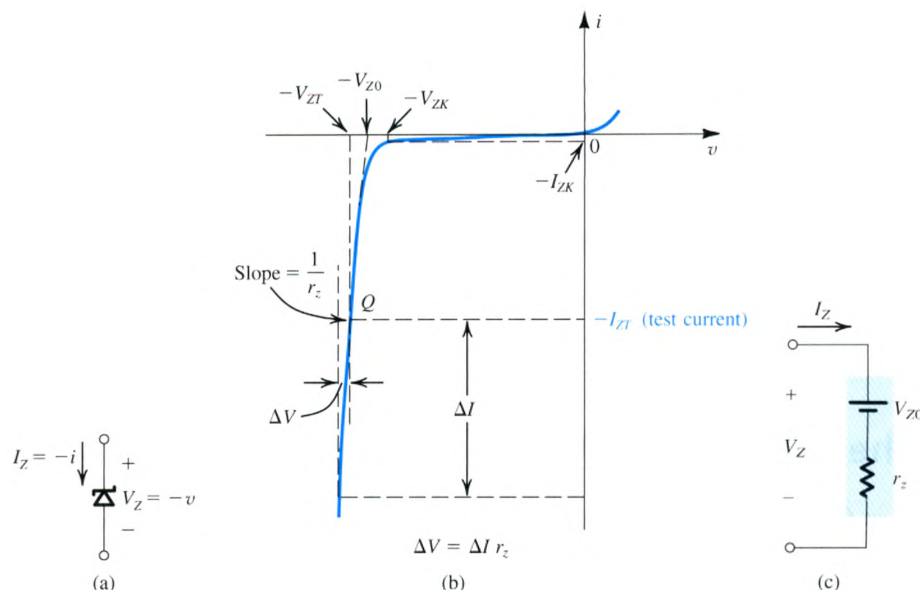


Figure 4.13 **(a)** Circuit symbol for the zener diode; **(b)** the diode i–v characteristic with the breakdown region shown in some detail; **(c)** model for the zener diode.

shows details of the diode i–v characteristic. The manufacturer will often specify several key parameters as indicated in Figure 4.13(b):

- The reverse voltage V_{ZT} at a test current I_{ZT} corresponding to the operating point Q. Typical values for V_{ZT} are below 10 V, although higher voltages are possible.
- The "knee" current I_{ZK}, corresponding to the reverse current at the onset of zener operation.
- The maximum power the device can safely dissipate. For example, a 0.5-W, 6.8-V zener diode can operate safely at currents up to a maximum of about 70 mA.
- The **temperature coefficient** (TC) of the zener voltage V_Z, commonly known as its **temco** and expressed in mV/°C.

The incremental resistance of the zener, r_z, will often also be specified. It relates changes in current, ΔI, to the resulting change in reverse voltage, $\Delta V = r_z \Delta I$. Note that r_z is inversely related to the slope of the i–v curve at Q in Fig. 4.13(b). Lower values of r_z indicate a steeper i–v curve and, hence, a more constant reverse voltage. The almost-linear i–v characteristic of the zener diode suggests it can be modeled as in Fig. 4.13(c). Here, V_{Z0} denotes the point at which the straight line of slope $1/r_z$ intersects the voltage axis in Fig. 4.13(b). The equivalent circuit model is another example of piecewise-linear modeling, and is analytically described by

$$V_Z = V_{Z0} + r_z I_Z \tag{4.8}$$

It is applicable in the vicinity of Q.

EXERCISES

4.15 The diode in Fig. 4.10 has a peak inverse voltage of 30 V. What is the minimum voltage V_{DD} that will ensure safe operation?
Ans. −30 V

4.16 A 200-mW 3.5-V zener diode is specified at a test current of 10 mA. Its incremental resistance is 10 Ω. What voltage will result when conducting a reverse current of 20 mA? What is the maximum current it can safely dissipate?
Ans. 3.6 V, 51 mA

4.4 The Small-Signal Model

We have so far considered the analysis of diode circuits using the nonlinear (exponential) model, and several piecewise-linear models including the constant-voltage-drop model, the ideal model, and the model for zener diodes in reverse breakdown. Linearized models may be solved directly and quickly, by hand or by a computer, but have so far provided only approximate results. In this section, we consider an alternative technique for modeling nonlinear devices that preserves the benefits of linear circuit analysis but provides more accuracy than piecewise-linear models as long as the circuit voltages and currents stay within a narrow range. These same methods will be applied to transistors in later chapters to enable rapid and intuitive analysis of amplifiers and other complex circuits.

Consider the situation in Fig. 4.14(a), where a dc voltage V_{DD} establishes a dc current I_D through the series combination of a resistance R and a diode D. The resulting diode voltage is denoted V_D. As mentioned above, we can find the values of I_D and V_D by solving the circuit using the diode exponential characteristic or, much more quickly, we can find approximate values using the diode constant-voltage-drop model.

Next, consider the situation of V_{DD} undergoing a small change ΔV_{DD}, as shown in Fig. 4.14(b). As indicated, the current I_D changes by an increment ΔI_D, and the diode voltage V_D changes by an increment ΔV_D. We want to find a quick way to determine the values of these incremental changes. Toward that end, we develop a "small-signal" model for the diode. Here the word *signal* emphasizes that in general, ΔV_{DD} can be a time-varying quantity. The qualifier "small" indicates that this diode model applies only when ΔV_D is kept sufficiently small, with "sufficiently" to be quantified shortly.

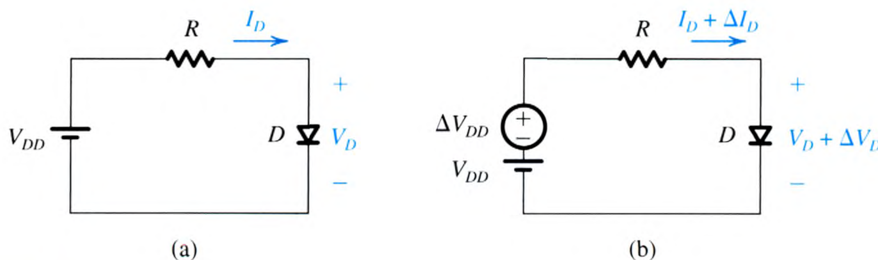


Figure 4.14 **(a)** A simple diode circuit; **(b)** the situation when V_{DD} changes by ΔV_{DD}.

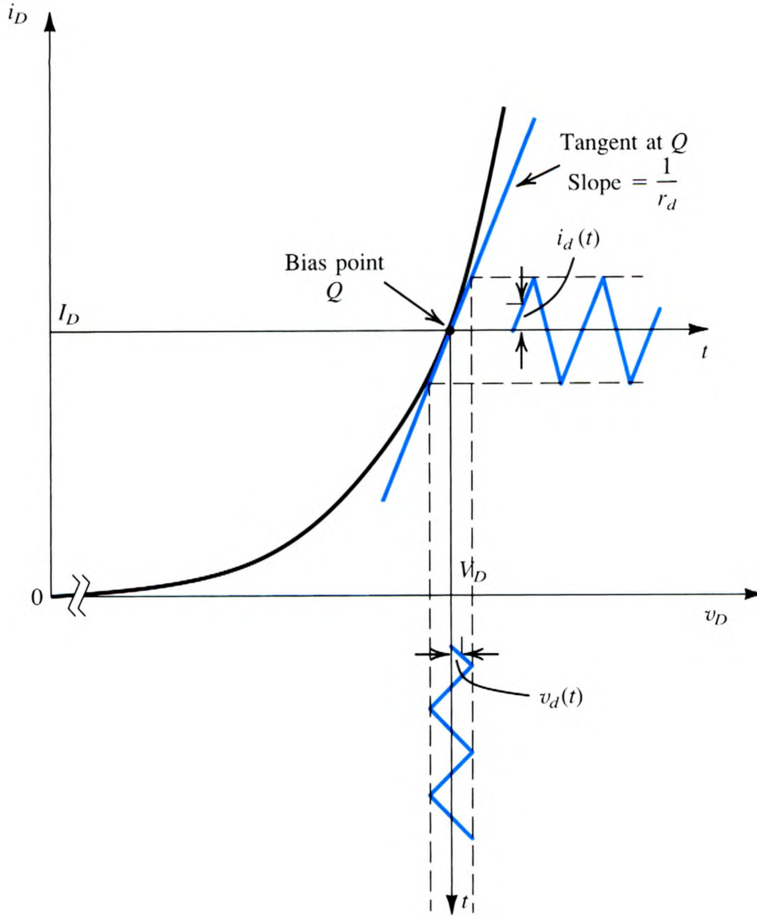


Figure 4.15 Development of the diode small-signal model.

To develop the diode small-signal model, refer to Fig. 4.15. We express the voltage across the diode as the sum of the dc voltage V_D and the time-varying signal $v_d(t)$,[5]

$$v_D(t) = V_D + v_d(t) \tag{4.9}$$

Correspondingly, the total instantaneous diode current $i_D(t)$ will be

$$i_D(t) = I_S e^{v_D/V_T} \tag{4.10}$$

Substituting for v_D from Eq. (4.9) gives

$$i_D(t) = I_S e^{(V_D + v_d)/V_T} = I_S e^{V_D/V_T} e^{v_d/V_T} \tag{4.11}$$

In the absence of the signal $v_d(t)$, the diode voltage is equal to V_D, and the diode current is

$$I_D = I_S e^{V_D/V_T} \tag{4.12}$$

[5]We hereafter use lowercase variables and subscripts, v_d and i_d, for small signals in place of ΔV_D and ΔI_D.

Thus, $i_D(t)$ in Eq. (4.11) can be expressed as

$$i_D(t) = I_D e^{v_d/V_T} \tag{4.13}$$

Now if the amplitude of the signal $v_d(t)$ is kept sufficiently small such that

$$\frac{v_d}{V_T} \ll 1 \tag{4.14}$$

then we may expand the exponential of Eq. (4.13) in a series and truncate the series after the first two terms to obtain the approximate expression

$$i_D(t) \simeq I_D\left(1 + \frac{v_d}{V_T}\right) \tag{4.15}$$

This is the **small-signal approximation**. It is valid when the variations in diode voltage, v_d, are kept smaller than about 5 mV (see Eq. 4.14, and recall that $V_T = 25$ mV).[6] Signal amplitudes elsewhere in the circuit may be much larger, so long as Eq. (4.14) is satisfied.

Thus, superimposed on the dc current I_D, we have a signal current component directly proportional to the signal voltage v_d. That is,

$$i_D = I_D + i_d \tag{4.16}$$

where, adopting the approximation in Eq. (4.15), we have

$$i_d = \frac{I_D}{V_T} v_d \Rightarrow v_d = \frac{V_T}{I_D} i_d \tag{4.17}$$

The quantity relating the signal current i_d to the signal voltage v_d is the **diode small-signal resistance**,

$$r_d = \frac{V_T}{I_D} \tag{4.18}$$

Note that the value of r_d is expressed in ohms and is inversely proportional to the bias current I_D.

We gain additional insight into the small-signal approximation and the small-signal diode model by considering again the graphical construction in Fig. 4.15. Here we see the diode operating at a dc bias point Q characterized by the dc voltage V_D and the corresponding dc current I_D. Superimposed on V_D we have a signal $v_d(t)$, assumed (arbitrarily) to have a triangular waveform.

It is easy to see that using the small-signal approximation is equivalent to assuming that *the signal amplitude is sufficiently small such that the excursion along the i–v curve is limited to a short almost-linear segment.* The slope of this segment is equal to the slope of the tangent to the i–v curve at the operating point Q. You are encouraged to prove that the slope of the i–v curve at $i = I_D$ is equal to I_D/V_T, which is $1/r_d$, the **small-signal conductance**. That is,

$$r_d = 1\bigg/\left[\frac{\partial i_D}{\partial v_D}\right]_{i_D = I_D} \tag{4.19}$$

[6]For $v_d = 5$ mV, $v_d/V_T = 0.2$. Thus the next term in the series expansion of the exponential will be $\frac{1}{2} \times 0.2^2 = 0.02$, a factor of 10 lower than the linear term we kept.

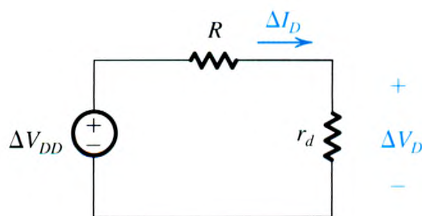


Figure 4.16 Circuit for determining the incremental quantities ΔI_D and ΔV_D for the circuit in Figure 4.14(b). Note that replacing the diode with its small-signal resistance r_d results in a linear circuit.

From the preceding we conclude that superimposed on the quantities V_D and I_D that define the dc bias point, or **quiescent point**, of the diode will be the small-signal quantities $v_d(t)$ and $i_d(t)$, which are related by the diode small-signal resistance r_d evaluated at the bias point (Eq. 4.18). Thus the small-signal analysis can be performed separately from the dc bias analysis, a great convenience that results from the linearization of the diode characteristics inherent in the small-signal approximation.

Specifically, we summarize the procedure as follows:

1. Perform dc analysis precisely using the exponential model, or approximately using the constant-voltage-drop model, or other piecewise-linear model. Thus, for the circuit in Fig. 4.14(b), we analyze the circuit in Fig. 4.14(a) to find I_D.
2. Linearize the circuit. For a forward-biased diode, we find r_d by substituting the current I_D into Eq. (4.18). We find the small-signal equivalent circuit by eliminating all independent dc sources (whose contribution to the final solution we already included in step 1) and replacing the diode with its small-signal resistance, r_d. For the circuit in Fig. 4.14(b), this results in the small-signal equivalent circuit in Fig. 4.16.
3. Solve the linearized circuit. For example, we find the incremental quantities ΔI_D and ΔV_D using the small-signal equivalent circuit in Fig. 4.16. We may want to check that the solution is consistent with our small-signal approximation. That is, $\Delta V_D \lesssim 5$ mV.

If necessary, we combine the results of the dc and small-signal analyses to find the complete solution. This is shown in Fig. 4.14(b), where the current is $I_D + \Delta I_D$.

We should not misinterpret the small-signal equivalent circuit. There is no actual diode resistance corresponding to r_d. It is merely a schematic representation of the linear relationship between the small-signal variables v_d and i_d. Nevertheless, small-signal equivalent circuits are essential to circuit analysis, providing intuition and fast, accurate solutions. We use these same steps, with different small-signal models, to analyze transistor circuits in Chapter 7.

Example 4.5

Consider the circuit shown in Fig. 4.17(a) for the case in which $R = 10\ \text{k}\Omega$. The power supply V^+ has a dc value of 10 V on which is superimposed a 60-Hz sinusoid of 1-V peak amplitude. (This "signal" component of the power-supply voltage is an imperfection in the power-supply design. It is known as the

Example 4.5 *continued*

power-supply ripple. More on this later.) Calculate both the dc voltage of the diode and the amplitude of the sine-wave signal appearing across it. Assume the diode to have a 0.7-V drop at 1-mA current.

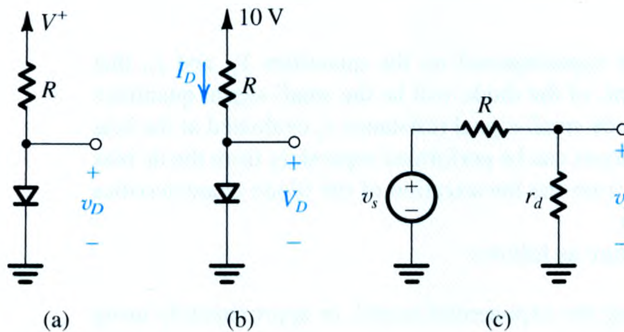


Figure 4.17 **(a)** Circuit for Example 4.5. **(b)** Circuit for calculating the dc operating point. **(c)** Small-signal equivalent circuit.

Solution

In step 1, we consider dc quantities only. We assume $V_D \simeq 0.7$ V and calculate the diode dc current

$$I_D = \frac{10 - 0.7}{10} = 0.93 \text{ mA}$$

Since this value is very close to 1 mA, the diode voltage will be very close to the assumed value of 0.7 V. In step 2, we linearize the circuit. The diode incremental resistance r_d is

$$r_d = \frac{V_T}{I_D} = \frac{25}{0.93} = 26.9\ \Omega$$

In step 3, we find the signal voltage across the diode from the small-signal equivalent circuit in Fig. 4.17(c). Here v_s denotes the 60-Hz 1-V peak sinusoidal component of V^+, and v_d is the corresponding signal across the diode. Using the voltage divider rule provides the peak amplitude of v_d as follows:

$$v_d(\text{peak}) = \hat{V}_s \frac{r_d}{R + r_d}$$

$$= 1 \frac{0.0269}{10 + 0.0269} = 2.68 \text{ mV}$$

Finally, we note that since this value is quite small, our use of the small-signal model of the diode is justified.

Finally, we note that while r_d models the small-signal operation of the diode at low frequencies, its dynamic operation is modeled by the capacitances C_j and C_d, which we studied in Section 3.6 and which also are small-signal parameters. A complete model of the diode includes C_j and C_d in parallel with r_d.

The diode models presented so far are summarized in Table 4.1 in the Summary Tables supplement found online at www.oup.com/he/sedra-smith8e.

4.5 Voltage Regulation

A **voltage regulator** is a circuit designed to provide a constant dc voltage between its output terminals. The output voltage must remain as constant as possible in spite of (a) changes in the load current drawn from the regulator output terminal and (b) changes in the dc power-supply voltage that feeds the regulator circuit. Since the forward-voltage drop of the diode remains almost constant at approximately 0.7 V while the current through it varies by relatively large amounts, a forward-biased diode can make a simple voltage regulator. For instance, we saw in Example 4.5 that while the 10-V dc supply voltage had a ripple of 2 V peak-to-peak (a ±10% variation), the corresponding ripple in the diode voltage was only about ±2.7 mV (a ±0.4% variation). We can obtain regulated voltages greater than 0.7 V by connecting a number of diodes in series. For example, three forward-biased diodes in series provide a voltage of about 2 V. We look at one such circuit in the following example, which utilizes the diode small-signal model to quantify the efficacy of the voltage regulator that is realized.

Example 4.6

Consider the circuit in Fig. 4.18. A string of three diodes is used to provide a constant voltage of about 2.1 V. We want to calculate the percentage change in this regulated voltage caused by (a) a ±10% change in the power-supply voltage, and (b) connection of a 1-kΩ load resistance.

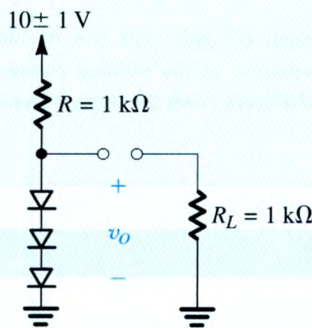


Figure 4.18 Circuit for Example 4.6.

Solution

With no load, the nominal value of the current in the diode string is given by

$$I = \frac{10-2.1}{1} = 7.9 \text{ mA}$$

Example 4.6 *continued*

Thus each diode will have an incremental resistance of

$$r_d = \frac{V_T}{I}$$

Thus,

$$r_d = \frac{25}{7.9} = 3.2\ \Omega$$

The three diodes in series will have a total incremental resistance of

$$r = 3r_d = 9.6\ \Omega$$

This resistance, along with the resistance R, forms a voltage divider whose ratio can be used to calculate the change in output voltage due to a $\pm 10\%$ (i.e., ± 1-V) change in supply voltage. Thus the peak-to-peak change in output voltage will be

$$\Delta v_O = 2\frac{r}{r+R} = 2\frac{0.0096}{0.0096+1} = 19 \text{ mV peak-to-peak}$$

That is, corresponding to the ± 1-V ($\pm 10\%$) change in supply voltage, the output voltage will change by ± 9.5 mV or $\pm 0.5\%$. Since this implies a change of about ± 3.2 mV per diode, our use of the small-signal model is justified.

When we connect a load resistance of 1 kΩ across the diode string, it draws a current of approximately 2.1 mA. Thus the current in the diodes decreases by 2.1 mA, resulting in a decrease in voltage across the diode string given by

$$\Delta v_O = -2.1 \times r = -2.1 \times 9.6 = -20 \text{ mV}$$

Since this implies that the voltage across each diode decreases by about 6.7 mV, our use of the small-signal model is not entirely justified. Nevertheless, a detailed calculation of the voltage change using the exponential model results in $\Delta v_O = -23$ mV, which is not too different from the approximate value we found using the incremental model.

EXERCISES

4.17 Find the value of the diode small-signal resistance r_d at bias currents of 0.1 mA, 1 mA, and 10 mA.
Ans. 250 Ω; 25 Ω; 2.5 Ω

4.18 Consider a diode biased at 1 mA. Find the change in current as a result of changing the voltage by (a) –10 mV, (b) –5 mV, (c) +5 mV, and (d) +10 mV. In each case, do the calculations (i) using the small-signal model and (ii) using the exponential model.
Ans. (a) –0.40, –0.33 mA; (b) –0.20, –0.18 mA; (c) +0.20, +0.22 mA; (d) +0.40, +0.49 mA

D4.19 Design the circuit of Fig. E4.19 so that $V_O = 3$ V when $I_L = 0$, and V_O changes by 20 mV per 1 mA of load current.

(a) Use the small-signal model of the diode to find the value of R.
(b) Specify the value of I_S of each of the diodes.
(c) For this design, use the diode exponential model to determine the actual change in V_O when a current $I_L = 1$ mA is drawn from the regulator.

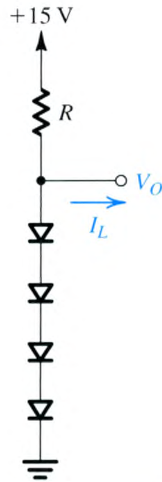


Figure E4.19

Ans. (a) $R = 2.4$ kΩ; (b) $I_S = 4.7 \times 10^{-16}$ A; (c) −23 mV.

Figure 4.18 is an example of a shunt regulator, so named because the regulator circuit appears in parallel (shunt) with the load.

We now illustrate, by way of an example, the use of zener diodes in the design of shunt regulators.

Example 4.7

The 6.8-V zener diode in the circuit of Fig. 4.19(a) is specified to have $V_Z = 6.8$ V at $I_Z = 5$ mA, $r_z = 20\ \Omega$, and $I_{ZK} = 0.2$ mA. The supply voltage V^+ is nominally 10 V but can vary by ±1 V.

(a) Find V_O with no load and with V^+ at its nominal value.
(b) Find the change in V_O resulting from the ±1-V change in V^+. Note that $(\Delta V_O/\Delta V^+)$, usually expressed in mV/V, is known as **line regulation**.
(c) Find the change in V_O resulting from connecting a load resistance R_L that draws a current $I_L = 1$ mA, and hence find the **load regulation** $(\Delta V_O/\Delta I_L)$ in mV/mA.
(d) Find the change in V_O when $R_L = 2$ kΩ.
(e) Find the value of V_O when $R_L = 0.5$ kΩ.
(f) What is the minimum value of R_L for which the diode still operates in the breakdown region?

Example 4.7 *continued*

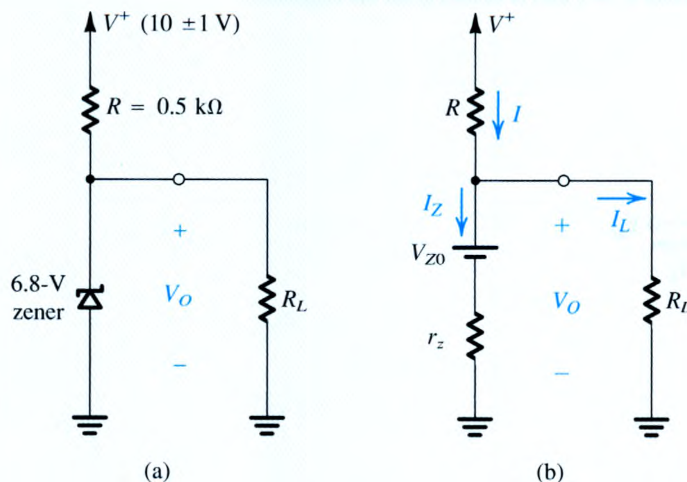


Figure 4.19 **(a)** Circuit for Example 4.7. **(b)** The circuit with the zener diode replaced with its equivalent circuit model.

Solution

First we must find the value of the parameter V_{Z0} of the zener diode model. (Refer to Fig. 4.13.) Substituting $V_Z = 6.8$ V, $I_Z = 5$ mA, and $r_z = 20\ \Omega$ in Eq. (4.8) yields $V_{Z0} = 6.7$ V. Figure 4.19(b) shows the circuit with the zener diode replaced with its model.

(a) With no load connected, the current through the zener is given by

$$I_Z = I = \frac{V^+ - V_{Z0}}{R + r_z} = \frac{10 - 6.7}{0.5 + 0.02} = 6.35 \text{ mA}$$

So,

$$V_O = V_{Z0} + I_Z r_z = 6.7 + 6.35 \times 0.02 = 6.83 \text{ V}$$

(b) For a ±1-V change in V^+, the change in output voltage can be found from

$$\Delta V_O = \Delta V^+ \frac{r_z}{R + r_z} = \pm 1 \times \frac{20}{500 + 20} = \pm 38.5 \text{ mV}$$

Thus,

$$\text{Line regulation} = 38.5 \text{ mV/V}$$

(c) When a load resistance R_L that draws a load current $I_L = 1$ mA is connected, the zener current will decrease by 1 mA. We find the corresponding change in zener voltage from

$$\Delta V_O = r_z \Delta I_Z = 20 \times -1 = -20 \text{ mV}$$

Example 4.7 *continued*

Thus the load regulation is

$$\text{Load regulation} \equiv \frac{\Delta V_O}{\Delta I_L} = -20 \text{ mV/mA}$$

(d) When a load resistance of 2 kΩ is connected, the load current will be approximately 6.8 V/2 kΩ = 3.4 mA. Thus the change in zener current will be $\Delta I_Z = -3.4$ mA, and the corresponding change in zener voltage (output voltage) will thus be

$$\Delta V_O = r_z \Delta I_Z$$
$$= 20 \times -3.4 = -68 \text{ mV}$$

This value could have been obtained by multiplying the load regulation by the value of I_L (3.4 mA).

(e) An R_L of 0.5 kΩ would draw a load current of 6.8/0.5 = 13.6 mA. This is not possible, because the current I supplied through R is only 6.4 mA (for $V^+ = 10$ V). Therefore, the zener must be cut off. If this is the case, then V_O is determined by the voltage divider formed by R_L and R (Fig. 4.19a),

$$V_O = V^+ \frac{R_L}{R + R_L}$$
$$= 10 \frac{0.5}{0.5 + 0.5} = 5 \text{ V}$$

Since this voltage is lower than the breakdown voltage of the zener, the diode is indeed no longer operating in the breakdown region.

(f) For the zener to be at the edge of the breakdown region, $I_Z = I_{ZK} = 0.2$ mA and $V_Z \simeq V_{ZK} \simeq 6.7$ V. At this point the lowest (worst-case) current supplied through R is $(9 - 6.7)/0.5 = 4.6$ mA, and thus the load current is $4.6 - 0.2 = 4.4$ mA. The corresponding value of R_L is

$$R_L = \frac{6.7}{4.4} \simeq 1.5 \text{ k}\Omega$$

Though simple and useful, shunt regulators based on diodes are inherently inefficient. Connecting the diode in parallel with the load means diode current is drawn in addition to load current, thereby dissipating extra power. Moreover, we can only produce the output voltages inherent to the diodes we select. Such circuits have been largely replaced by specially designed integrated circuits that perform voltage regulation more efficiently and with greater flexibility.

EXERCISE

4.20 A zener diode whose nominal voltage is 6 V at 5 mA has an incremental resistance of 80 Ω. What voltage do you expect if the diode current is halved? Doubled? What is the value of V_{Z0} in the zener model?

Ans. 5.8 V; 6.4 V; 5.6 V

4.6 Rectifier Circuits

An important application of diodes is in the design of rectifier circuits. A diode rectifier forms an essential building block of the dc power supplies required to power electronic equipment. Figure 4.20 shows a block diagram of such a power supply. As indicated, the power supply is fed from the 120-V (rms)[7] 60-Hz ac line, and it delivers a dc voltage V_O (usually in the range of 3 V to 20 V) to an electronic circuit represented by the *load* block. The dc voltage V_O is required to be as constant as possible in spite of variations in the ac line voltage and in the current drawn by the load.

The first block in a dc power supply is the **power transformer**. It consists of two separate coils wound around an iron core that magnetically couples the two windings. The **primary winding**, having N_1 turns, is connected to the 120-V ac supply, and the **secondary winding**, having N_2 turns, is connected to the circuit of the dc power supply. Thus an ac voltage v_S of $120(N_2/N_1)$ V (rms) develops between the two terminals of the secondary winding. By selecting an appropriate turns ratio (N_1/N_2) for the transformer, the designer can step the line voltage down to the value required to yield the particular dc voltage output of the supply. For instance, a secondary voltage of 8-V rms may be appropriate for a dc output of 5 V. We can achieve this with a 15:1 turns ratio.

In addition to providing the appropriate sinusoidal amplitude for the dc power supply, the power transformer provides electrical isolation between the electronic equipment and the power-line circuit. This isolation minimizes the risk of electric shock to the equipment user.

The diode rectifier converts the input sinusoid v_S to a unipolar output, which can have the pulsating waveform indicated in Fig. 4.20. Although this waveform has a nonzero average or a dc component, its pulsating nature makes it unsuitable as a dc source for electronic circuits, hence the need for a filter. The variations in the magnitude of the rectifier output are considerably reduced by the filter block in Fig. 4.20. In this section we'll look at a number of rectifier circuits and a simple implementation of the output filter.

The output of the rectifier filter, though much more constant than without the filter, still contains a time-dependent component, known as **ripple**. To reduce the ripple and to stabilize the magnitude of the dc output voltage against variations caused by changes in load current, we use a voltage regulator. It can be a shunt regulator like those studied in Section 4.5. Alternatively, and much more commonly these days, we use an integrated-circuit regulator.

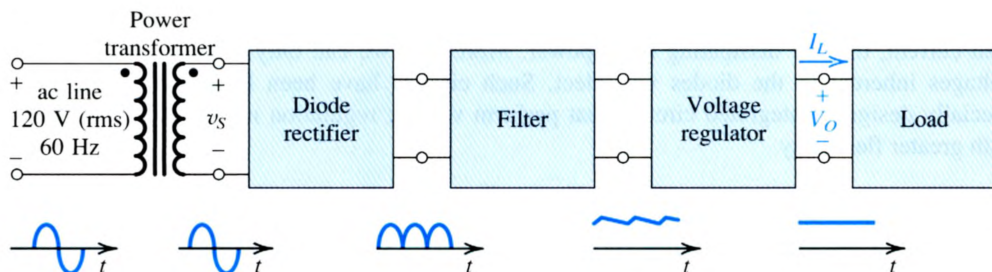


Figure 4.20 Block diagram of a dc power supply.

[7]A reminder that ac voltages are typically specified by their root mean square (rms) value.

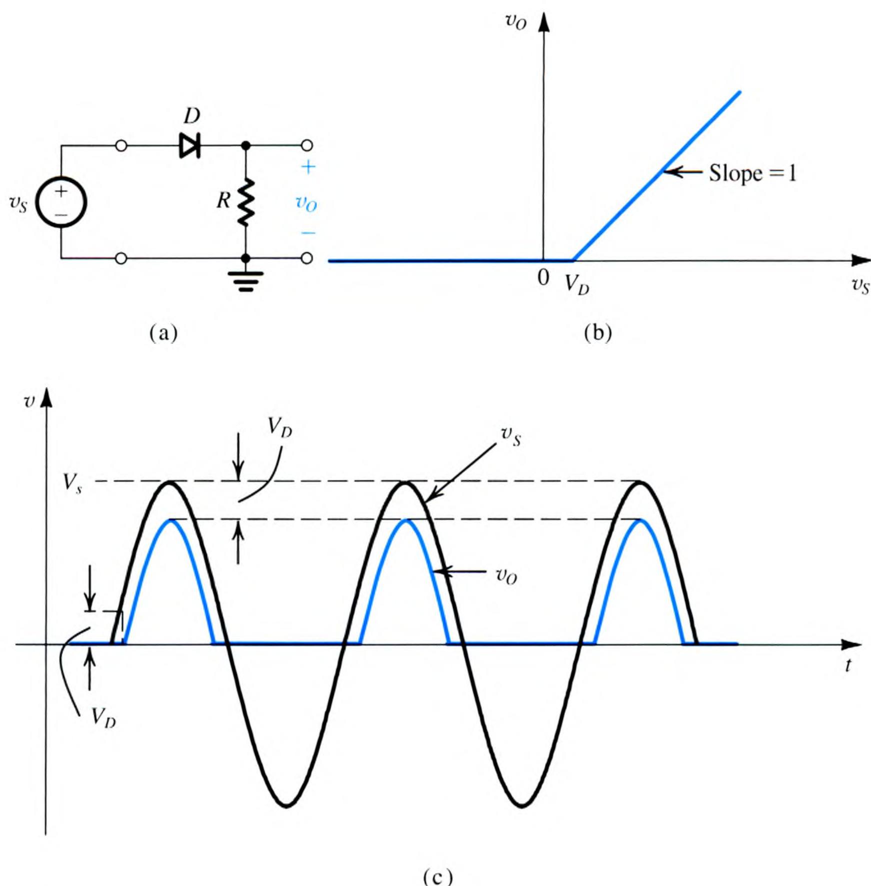


Figure 4.21 **(a)** Half-wave rectifier. **(b)** Transfer characteristic of the rectifier circuit. **(c)** Input and output waveforms.

4.6.1 The Half-Wave Rectifier

The half-wave rectifier uses alternate half-cycles of the input sinusoid. Figure 4.21(a) shows the circuit of a half-wave rectifier. We analyzed this circuit in Section 4.1 (see Fig. 4.3) assuming an ideal diode. Using the more realistic constant-voltage-drop diode model,

$$v_O = 0, \qquad v_S < V_D \tag{4.20a}$$

$$v_O = v_S - V_D, \qquad v_S \geq V_D \tag{4.20b}$$

We have sketched the transfer characteristic represented by these equations in Fig. 4.21(b). We continue to assume a forward voltage $V_D = 0.7$ V, although rectifier diodes conducting large forward current may have $V_D = 0.8$ V to 1.0 V. Figure 4.21(c) shows the output voltage obtained when the input v_S is a sinusoid.

In selecting diodes for rectifier design, we specify two important parameters: the current-handling capability required of the diode, determined by the largest forward current the diode is expected to conduct, and the **peak inverse voltage** (PIV) that the diode must be able to withstand without breakdown, determined by the largest reverse voltage that is expected to appear across the diode. In the rectifier circuit of Fig. 4.21(a), we observe that

when v_S is negative, the diode will be cut off and v_O will be zero. It follows that the PIV is equal to the peak of v_S,

$$\text{PIV} = V_s \tag{4.21}$$

It is usually a good idea, however, to select a diode that has a reverse breakdown voltage at least 50% greater than the expected PIV.

Before leaving the half-wave rectifier, note two points. First, it is possible to use the diode exponential characteristic to determine the exact transfer characteristic of the rectifier (see Problem 4.66). However, the amount of work involved is usually too great to be justified in practice. Of course, we can easily do the analysis using a computer circuit-analysis program such as SPICE.

Second, whether we analyze the circuit accurately or not, it should be obvious that this circuit does not function properly when the input signal is small. For instance, this circuit cannot be used to rectify an input sinusoid of 100-mV amplitude. For such an application we resort to a so-called precision rectifier, which uses diodes in conjunction with op amps. One such circuit is presented in Section 4.6.5.

EXERCISE

4.21 For the half-wave rectifier circuit in Fig. 4.21(a), show the following: (a) For the half-cycles during which the diode conducts, conduction begins at an angle $\theta = \sin^{-1}(V_D/V_s)$ and terminates at $(\pi - \theta)$, for a total conduction angle of $(\pi - 2\theta)$. (b) The average value (dc component) of v_O is $V_O \simeq (1/\pi)V_s - V_D/2$. (c) The peak diode current is $(V_s - V_D)/R$.
Find numerical values for these quantities for the case of 12-V (rms) sinusoidal input, $V_D \simeq 0.7$ V, and $R = 100\ \Omega$. Also, give the value for PIV.
Ans. (a) $\theta = 2.4°$, conduction angle = 175°; (b) 5.05 V; (c) 163 mA; 17 V

4.6.2 The Full-Wave Rectifier

The full-wave rectifier uses both halves of the input sinusoid. To provide a unipolar output, it inverts the negative halves of the sine wave. One possible implementation is shown in Fig. 4.22(a). Here the transformer secondary winding is **center-tapped** to provide two equal voltages v_S across the two halves of the secondary winding with the polarities indicated. Note that when the input line voltage (feeding the primary) is positive, both of the signals labeled v_S will be positive. In this case D_1 will conduct and D_2 will be reverse biased. The current through D_1 will flow through R and back to the center tap of the secondary. The circuit then behaves like a half-wave rectifier, and the output during the positive half-cycles when D_1 conducts will be identical to that produced by the half-wave rectifier.

Now, during the negative half-cycle of the ac line voltage, both of the voltages labeled v_S will be negative. Thus D_1 will be cut off while D_2 will conduct. The current conducted by D_2 will flow through R and back to the center tap. It follows that during the negative half-cycles while D_2 conducts, the circuit behaves again as a half-wave rectifier. The important point,

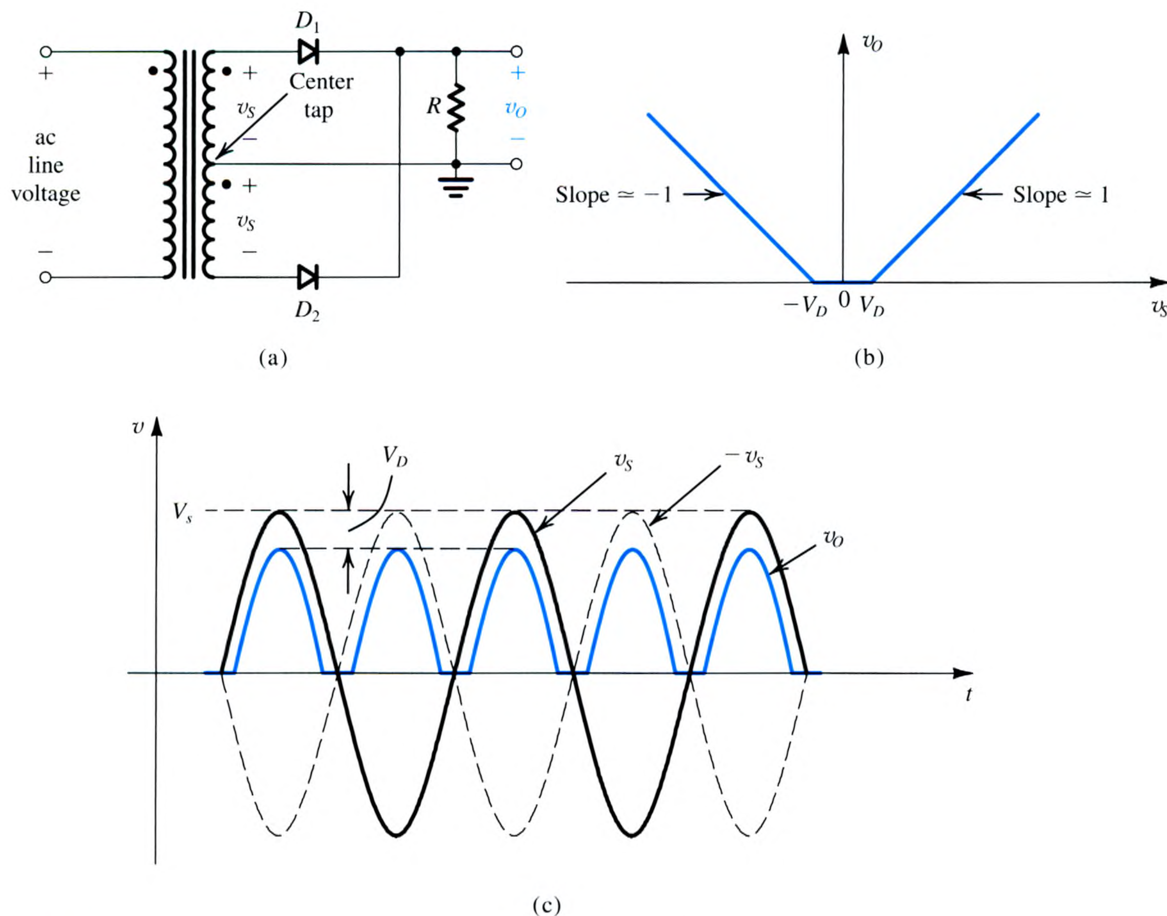


Figure 4.22 Full-wave rectifier using a transformer with a center-tapped secondary winding: **(a)** circuit; **(b)** transfer characteristic assuming a constant-voltage-drop model for the diodes; **(c)** input and output waveforms.

however, is that the current through R always flows in the same direction, and thus v_O will be unipolar, as indicated in Fig. 4.22(c). We obtain the output waveform again using the diode model with constant forward voltage drop, V_D. The resulting transfer characteristic is shown in Fig. 4.22(b).

The full-wave rectifier obviously produces a more "energetic" waveform than the half-wave rectifier. In almost all rectifier applications, we opt for a full-wave type of some kind.

To find the PIV of the diodes in the full-wave rectifier circuit, consider the situation during the positive half-cycles. Diode D_1 is conducting, and D_2 is cut off. The voltage at the cathode of D_2 is v_O, and that at its anode is $-v_S$. Thus the reverse voltage across D_2 will be $(v_O + v_S)$, which will reach its maximum when v_O is at its peak value of $(V_s - V_D)$, and v_S is at its peak value of V_s; thus,

$$\text{PIV} = 2V_s - V_D$$

which is approximately twice that for the case of the half-wave rectifier.

Video Example VE 4.4 Analysis of a Full-Wave Rectifier

The full-wave rectifier circuit in Fig. 4.22(a) with a 1-kΩ load operates from a 120-V (rms) 60-Hz household supply through a 6-to-1 transformer. It uses two silicon diodes that can be modeled to have a 0.7-V drop for all currents. What is the peak voltage of the rectified output? For what fraction of a cycle does each diode conduct? What is the average output voltage? What is the average current in the load?


VE 4.4

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: **4.69**

EXERCISE

4.22 For the full-wave rectifier circuit in Fig. 4.22(a), show the following: (a) The output is zero for an angle of $2\sin^{-1}(V_D/V_s)$ centered around the zero-crossing points of the sine-wave input. (b) The average value (dc component) of v_O is $V_O \simeq (2/\pi)V_s - V_D$. (c) The peak current through each diode is $(V_s - V_D)/R$. Find the fraction (percentage) of each cycle during which $v_O > 0$, the value of V_O, the peak diode current, and the value of PIV, all for the case in which v_S is a 12-V (rms) sinusoid, $V_D \simeq 0.7$ V, and $R = 100\ \Omega$.
Ans. 97.4%; 10.1 V; 163 mA; 33.2 V

4.6.3 The Bridge Rectifier

An alternative implementation of the full-wave rectifier is shown in Fig. 4.23(a). This circuit, known as the bridge rectifier because of the similarity of its configuration to that of the Wheatstone bridge, does not require a center-tapped transformer, giving it a distinct advantage over the full-wave rectifier circuit of Fig. 4.22. The bridge rectifier, however, requires four diodes as compared to two in the previous circuit. This is not much of a disadvantage, because diodes are inexpensive and we can buy a diode bridge in one package.

The bridge-rectifier circuit operates like this: During the positive half-cycles of the input voltage, v_S is positive, and so current is conducted through diode D_1, resistor R, and diode D_2, as in Fig 4.23(b). Meanwhile, diodes D_3 and D_4 will be reverse biased. Notice that there are two diodes in series in the conduction path, and so v_O will be lower than v_S by two diode drops (compared to one drop in the circuit previously discussed). This is somewhat of a disadvantage of the bridge rectifier.

Next, consider the situation during the negative half-cycles of the input voltage, shown in Fig. 4.23(c). The secondary voltage v_S will be negative, so $-v_S$ will be positive, forcing current through D_3, R, and D_4. Meanwhile, diodes D_1 and D_2 will be reverse biased. The

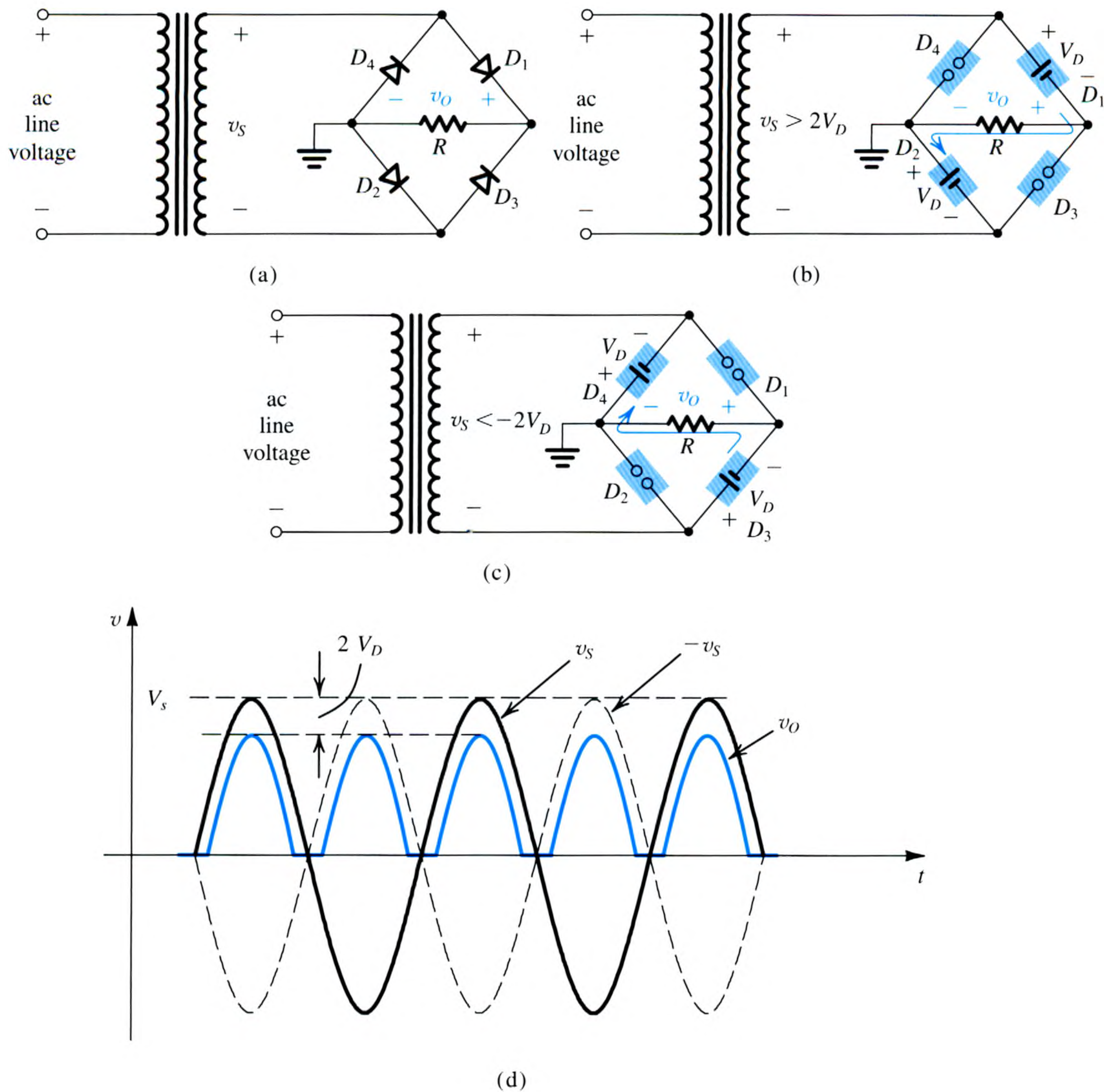


Figure 4.23 The bridge rectifier: **(a)** circuit; **(b)** the equivalent circuit during positive half-cycles; **(c)** the equivalent circuit during negative half-cycles; **(d)** input and output waveforms.

important point to note, though, is that during both half-cycles, current flows through R in the same direction (from right to left), and thus v_O will always be positive, as indicated in Fig. 4.23(d).

To determine the peak inverse voltage (PIV) of each diode, consider the circuit during the positive half-cycles. The reverse voltage across D_3 can be found from the loop formed by D_3, R, and D_2 as

$$v_{D3}(\text{reverse}) = v_O + v_{D2}(\text{forward})$$

Thus the maximum value of v_{D3} occurs at the peak of v_O and is given by

$$\text{PIV} = V_s - 2V_D + V_D = V_s - V_D$$

Notice that here the PIV is about half the value for the full-wave rectifier with a center-tapped transformer. This is another advantage of the bridge rectifier.

Yet one more advantage of the bridge-rectifier circuit over one using a center-tapped transformer is that the secondary winding only needs about half as many turns. Another way of looking at this is to notice that each half of the secondary winding of the center-tapped transformer is used for only half the time. These advantages have made the bridge rectifier the most popular rectifier circuit configuration.

EXERCISE

4.23 For the bridge-rectifier circuit of Fig. 4.23(a), use the constant-voltage-drop diode model to show that (a) the average (or dc component) of the output voltage is $V_O \simeq (2/\pi)V_s - 2V_D$ and (b) the peak diode current is $(V_s - 2V_D)/R$. Find numerical values for the quantities in (a) and (b) and the PIV for the case in which v_S is a 12-V (rms) sinusoid, $V_D \simeq 0.7$ V, and $R = 100\ \Omega$.
Ans. 9.4 V; 156 mA; 16.3 V

4.6.4 The Rectifier with a Filter Capacitor—The Peak Rectifier

The pulsating nature of the output voltage produced by the rectifier circuits discussed above makes it unsuitable as a dc supply for electronic circuits. A simple way to reduce the variation of the output voltage is to place a capacitor across the load resistor. It will be shown that this **filter capacitor** substantially reduces the variations in the rectifier output voltage.

To see how the rectifier circuit with a filter capacitor works, consider first the simple circuit shown in Fig. 4.24. Let the input v_I be a sinusoid with a peak value V_p, and assume the diode to be ideal. As v_I goes positive, the diode conducts and the capacitor is charged so that $v_O = v_I$. This situation continues until v_I reaches its peak value V_p. Beyond the peak, as v_I decreases, the diode becomes reverse biased and the output voltage remains constant at the value V_p. In fact, theoretically speaking, the capacitor will retain its charge and hence its voltage indefinitely, because there is no way for the capacitor to discharge. So the circuit provides a dc voltage output equal to the peak of the input sine wave. This is a very encouraging result since we want to produce a dc output.

Next, consider the more practical situation where a load resistance R is connected across the capacitor C, as shown in Fig. 4.25(a). We will continue to assume the diode to be ideal. As before, for a sinusoidal input, the capacitor charges to the peak of the input V_p. Then the diode cuts off, and the capacitor discharges through the load resistance R. The capacitor discharge will continue for almost the entire cycle, until the time at which v_I exceeds the capacitor voltage. Then the diode turns on again and charges the capacitor up to the peak of v_I, and the process repeats itself. Notice that to keep the output voltage from decreasing too much during capacitor discharge, we select a value for C so that the time constant CR is much greater than the discharge interval.

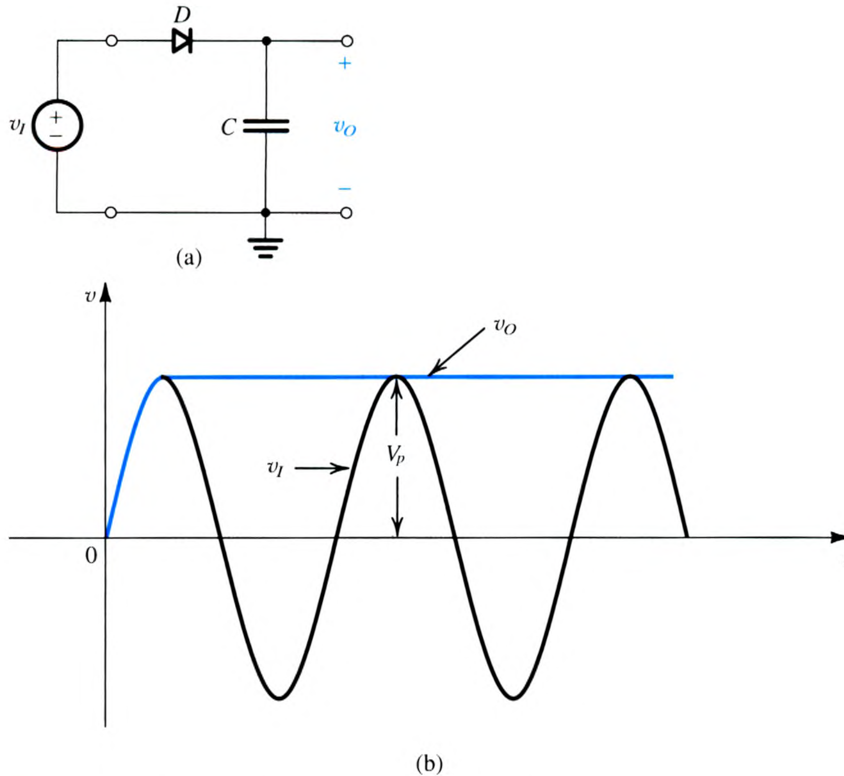


Figure 4.24 **(a)** A simple circuit used to illustrate the effect of a filter capacitor. **(b)** Input and output waveforms assuming an ideal diode. Note that the circuit provides a dc voltage equal to the peak of the input sine wave. The circuit is therefore known as a *peak rectifier* or a *peak detector*.

We are now ready to analyze the circuit in detail. Figure 4.25(b) shows the steady-state input and output voltage waveforms under the assumption that $CR \gg T$, where T is the period of the input sinusoid. The waveforms of the load current

$$i_L = v_O/R \tag{4.22}$$

and of the diode current (when it is conducting)

$$i_D = i_C + i_L = C\frac{dv_I}{dt} + i_L \tag{4.23}$$

are shown in Fig. 4.25(c). The following observations are in order:

1. The diode conducts for a brief interval, Δt, near the peak of the input sinusoid and supplies the capacitor with charge equal to that lost during the much longer discharge interval. The latter is approximately equal to the period T.
2. Assuming an ideal diode, the diode conduction begins at time t_1, at which the input v_I equals the exponentially decaying output v_O. Conduction stops at t_2 shortly after the peak of v_I; we can determine the exact value of t_2 by setting $i_D = 0$ in Eq. (4.23).
3. During the diode-off interval, the capacitor C discharges through R, and so v_O decays exponentially with a time constant CR. The discharge interval begins just past the

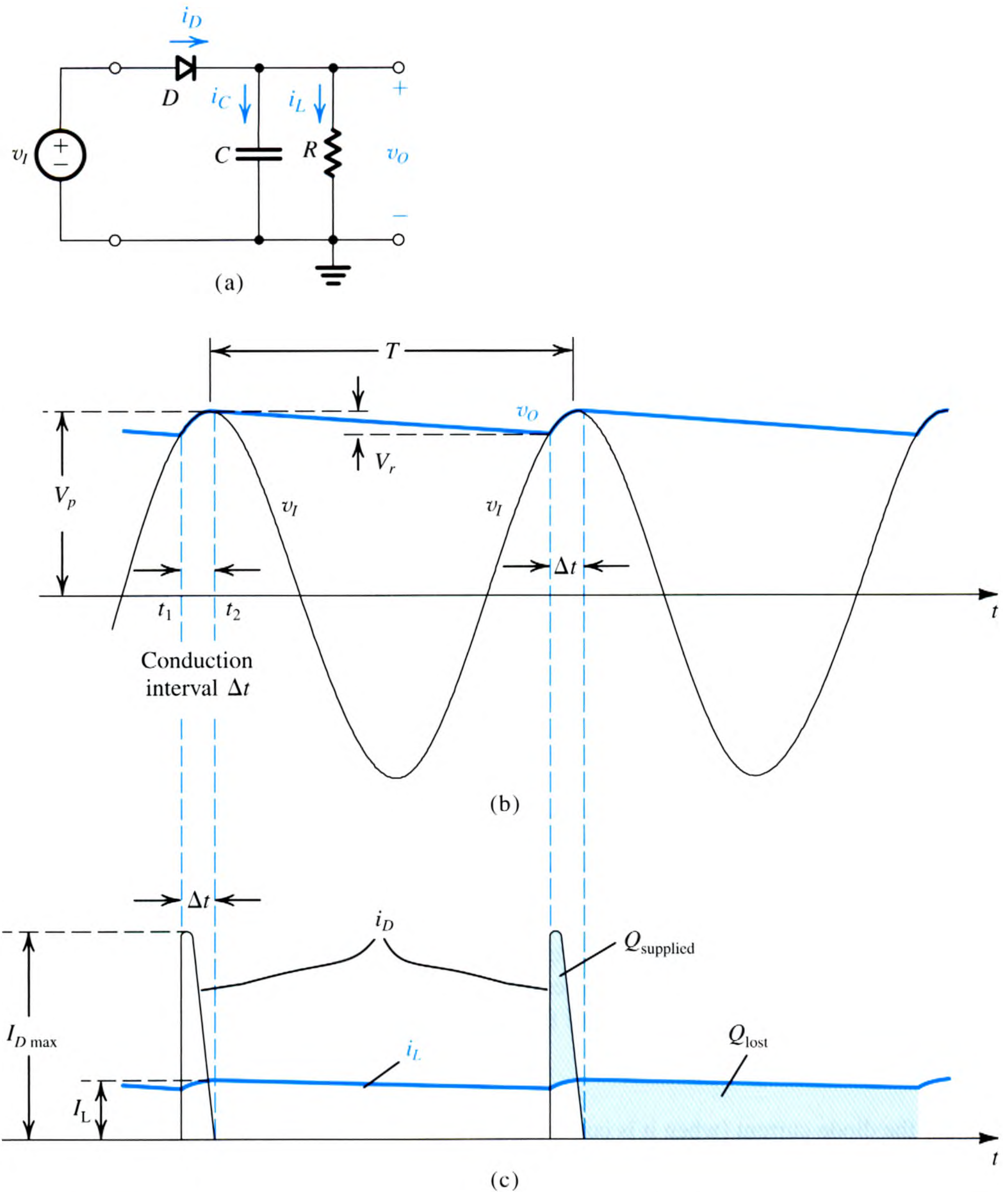


Figure 4.25 Voltage and current waveforms in the peak-rectifier circuit with $CR \gg T$. The diode is assumed ideal.

peak of v_I. At the end of the discharge interval, which lasts for almost the entire period T, $v_O = V_p - V_r$, where V_r is the peak-to-peak ripple voltage. When $CR \gg T$, the value of V_r is small.

4. When V_r is small, v_O is almost constant and equal to the peak value of v_I. Thus the dc output voltage is approximately equal to V_p. Similarly, the current i_L is almost

constant, and its dc component I_L is given by

$$I_L = \frac{V_p}{R} \tag{4.24}$$

If needed, we can derive a more accurate expression for the output dc voltage by taking the average of the extreme values of v_O,

$$V_O = V_p - \tfrac{1}{2}V_r \tag{4.25}$$

This allows us to derive expressions for V_r and for the average and peak values of the diode current. During the diode-off interval, v_O can be expressed as

$$v_O = V_p e^{-t/CR}$$

At the end of the discharge interval we have

$$V_p - V_r \simeq V_p e^{-T/CR}$$

Now, since $CR \gg T$, we can use the approximation $e^{-T/CR} \simeq 1 - T/CR$ to get

$$V_r \simeq V_p \frac{T}{CR} \tag{4.26}$$

To keep V_r small we must select a capacitance C so that $CR \gg T$. The **ripple voltage** V_r in Eq. (4.26) can be expressed in terms of the frequency $f = 1/T$ as

$$V_r = \frac{V_p}{fCR} \tag{4.27a}$$

Using Eq. (4.24) we can express V_r by the alternate expression

$$V_r = \frac{I_L}{fC} \tag{4.27b}$$

Note that an alternative interpretation of the approximation made in Eq. (4.27b) is that the capacitor discharges by means of a constant current $I_L = V_p/R$. This approximation is valid as long as $V_r \ll V_p$.

Assuming that diode conduction ceases almost at the peak of v_I, we can determine the **conduction interval** Δt from

$$V_p \cos(\omega \Delta t) = V_p - V_r$$

where $\omega = 2\pi f = 2\pi/T$ is the angular frequency of v_I. Since $(\omega \Delta t)$ is a small angle, we can use the approximation $\cos(\omega \Delta t) \simeq 1 - \frac{1}{2}(\omega \Delta t)^2$ to obtain

$$\omega \Delta t = \sqrt{2V_r/V_p} \implies \Delta t = \frac{T}{2\pi}\sqrt{\frac{2V_r}{V_p}} \tag{4.28}$$

We note that when $V_r \ll V_p$, the conduction angle $\omega \Delta t$ will be small, as we assumed.

Another useful interpretation is that the charge supplied by the diode to the capacitor during the conduction interval must equal the charge that the capacitor loses during the

discharge interval. The average capacitor, diode, and load currents during conduction are related by Eq. (4.23), $i_{Cav} = i_{Dav} - I_L$. Hence, the supplied charge is

$$Q_{\text{supplied}} = i_{Cav}\,\Delta t \simeq (i_{Dav} - I_L)\frac{T}{2\pi}\sqrt{\frac{2V_r}{V_p}} \tag{4.29}$$

and the lost charge is

$$Q_{\text{lost}} = CV_r = I_L T \tag{4.30}$$

These charges are given by the shaded areas of $(i_D - i_L)$ on Fig. 4.25(c), which must be equal. Equating Eqs. (4.29) and (4.30), the average diode current during conduction is

$$i_{Dav} = I_L\left(1 + \pi\sqrt{2V_p/V_r}\right) \tag{4.31}$$

To select an appropriate diode, we must find the peak value of the diode current. We do this by evaluating Eq. (4.23) at the onset of conduction ($t = -\Delta t$ where $t = 0$ is at the peak) and using Eq. (4.24):

$$i_{Dmax} = I_L\left(1 + 2\pi\sqrt{2V_p/V_r}\right) \tag{4.32}$$

From Eqs. (4.31) and (4.32), we see that for $V_r \ll V_p$, $i_{Dmax} \simeq 2i_{Dav}$, which correlates with the fact that the waveform of i_D is almost a right-angle triangle (see Fig. 4.25c).

Example 4.8

Consider a peak rectifier fed by a 60-Hz sinusoid with a peak value $V_p = 100$ V. Let the load resistance $R = 10$ kΩ. Find the value of the capacitance C that will result in a peak-to-peak ripple of 2 V. Also, calculate the fraction of the cycle during which the diode is conducting and the average and peak values of the diode current.

Solution

From Eq. (4.27a) we obtain the value of C as

$$C = \frac{V_p}{V_r fR} = \frac{100}{2 \times 60 \times 10 \times 10^3} = 83.3\ \mu\text{F}$$

The conduction angle $\omega\Delta t$ is found from Eq. (4.28) as

$$\omega\Delta t = \sqrt{2 \times 2/100} = 0.2 \text{ rad}$$

Thus the diode conducts for $(0.2/2\pi) \times 100 = 3.18\%$ of the cycle. The average diode current is obtained from Eq. (4.31), where $I_L = 100/10 = 10$ mA, as

$$i_{Dav} = 10\left(1 + \pi\sqrt{2 \times 100/2}\right) = 324 \text{ mA}$$

We find the peak diode current using Eq. (4.32),

$$i_{Dmax} = 10\left(1 + 2\pi\sqrt{2 \times 100/2}\right) = 638 \text{ mA}$$

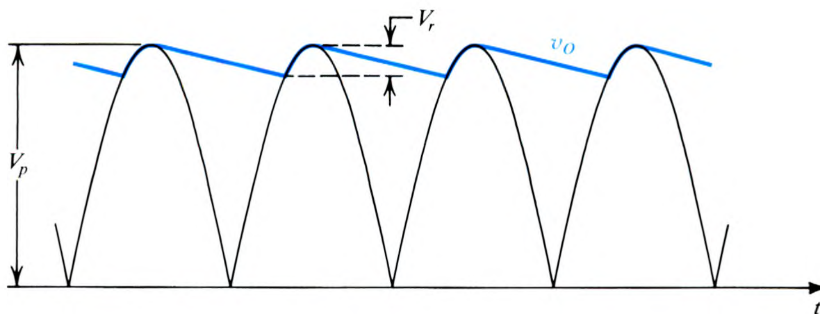


Figure 4.26 Waveforms in the full-wave peak rectifier.

The circuit of Fig. 4.25(a) is known as a half-wave **peak rectifier**. The full-wave rectifier circuits of Figs. 4.22(a) and 4.23(a) can be converted to peak rectifiers by including a capacitor across the load resistor. As in the half-wave case, the output dc voltage will be almost equal to the peak value of the input sine wave (Fig. 4.26). The ripple frequency, however, will be twice that of the input. We can find the peak-to-peak ripple voltage for this case using a procedure identical to that above but with the discharge period T replaced by $T/2$, resulting in

$$V_r = \frac{V_p}{2fCR} \tag{4.33}$$

While the diode conduction interval, Δt, will still be given by Eq. (4.28), the average and peak currents in each of the diodes will be given by

$$i_{Dav} = I_L\left(1 + \pi\sqrt{V_p/2V_r}\right) \tag{4.34}$$

$$i_{Dmax} = I_L\left(1 + 2\pi\sqrt{V_p/2V_r}\right) \tag{4.35}$$

Comparing these expressions with the corresponding ones for the half-wave case, we note that for the same values of V_p, f, R, and V_r (and thus the same I_L), we need a capacitor half the size of that required in the half-wave rectifier. Also, the current in each diode in the full-wave rectifier is approximately half the current flowing in the diode of the half-wave circuit.

The analysis above assumed ideal diodes. We can improve the accuracy of the results by taking the diode voltage drop into account. We can easily do this by replacing the peak voltage V_p to which the capacitor charges with $(V_p - V_D)$ for the half-wave circuit and the full-wave circuit using a center-tapped transformer and with $(V_p - 2V_D)$ for the bridge-rectifier case.

We conclude this section by noting that peak-rectifier circuits find application in signal-processing systems where it is required to detect the peak of an input signal. In such a case, the circuit is referred to as a **peak detector**. A particularly popular application of the peak detector is in the design of a demodulator for amplitude-modulated (AM) signals. AM signals are useful when a low-bandwidth signal is to be communicated using a high-frequency portion of spectrum. For example, AM radios receive audio signals (where

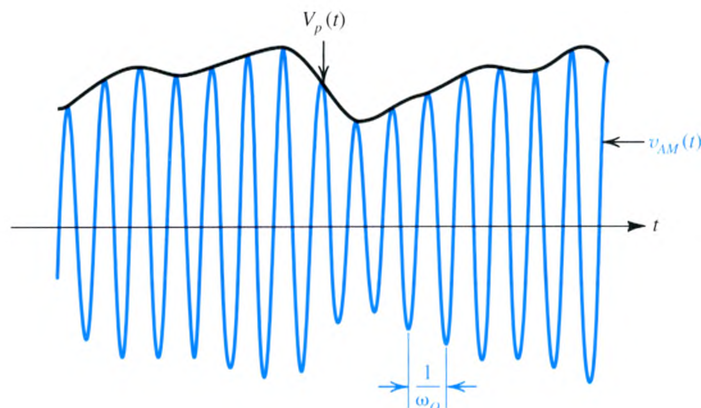


Figure 4.27 The amplitude modulated (AM) signal v_{AM} is a sinusoid at a frequency ω_o, called the carrier frequency, and has an amplitude modulated by the information-bearing signal V_p.

bandwidth is limited to less than 10 kHz) that are communicated using spectrum around 1MHz.

AM signals are sinusoids with a time-varying amplitude, $V_p(t)$, that bears the information being communicated (e.g., the audio signal in an AM radio).

$$v_{AM}(t) = V_p(t)\sin(\omega_0 t)$$

Typically, the variations in $V_p(t)$ occur much more slowly than the underlying sinusoid (called a **carrier**) at ω_0, so that $V_p(t)$ is visible by looking at the "envelope" of the waveform, as illustrated in Fig. 4.27. The output of a peak detector will approximately follow this envelope, and is thus useful as an AM demodulator. To avoid excessive drop of the output between successive peaks, we must ensure $1/CR < \omega_0$. On the other hand, if CR is too large, the peak detector output will not be able to follow declines in $V_p(t)$, so $1/CR$ must be larger than the highest frequency components of $V_p(t)$.

EXERCISES

4.24 Derive the expressions in Eqs. (4.33), (4.34), and (4.35).

4.25 Consider a bridge-rectifier circuit with a filter capacitor C placed across the load resistor R for the case in which the transformer secondary delivers a sinusoid of 12 V (rms) having a 60-Hz frequency and assuming $V_D = 0.8$ V and a load resistance $R = 100\ \Omega$. Find the value of C that results in a ripple voltage no larger than 1 V peak-to-peak. What is the dc voltage at the output? Find the load current. Find the diodes' conduction angle. Provide the average and peak diode currents. What is the peak reverse voltage across each diode? Specify the diode in terms of its peak current and its PIV.
Ans. 1281 μF; 15.4 V or (a better estimate) 14.9 V; 0.15 A; 0.36 rad (20.7°); 1.45 A; 2.74 A; 16.2 V. Thus select a diode with 3.5-A to 4-A peak current and a 20-V PIV rating.

THE EARLIEST SEMICONDUCTOR DIODE

The cat's whisker, or crystal detector, was the first electronic diode to be commercialized as an envelope detector for the radio-frequency signals used in radio telephony. The earliest diode, invented in Germany by Karl Ferdinand Braun, consisted of a small slab of galena (lead sulfide) to which contact was made by sharpened spring wire, which could be adjusted. For this and other contributions to early radios, Braun received the Nobel Prize in Physics in 1909. The silicon-based point-contact diode, later refined and packaged, was an important solid-state component of radar equipment during World War II.

4.6.5 Precision Half-Wave Rectifier—The Superdiode[8]

The rectifier circuits we have studied so far suffer from having one or two diode drops in the signal paths. Thus these circuits work well only when the signal to be rectified is much larger than the voltage drop of a conducting diode (0.7 V or so). In such a case, the details of the diode forward characteristics or the exact value of the diode voltage do not play a prominent role in determining circuit performance. This is indeed the case in the application of rectifier circuits in power-supply design. There are other applications, however, where the signal to be rectified is small (e.g., on the order of 100 mV or so) and thus clearly insufficient to turn on a diode. Also, in instrumentation applications, we often need rectifier circuits with very precise and predictable transfer characteristics. For these applications, a class of circuits has been developed using op amps (Chapter 2) together with diodes to provide precision rectification. In the following discussion, we study one such circuit. A comprehensive study of op amp–diode circuits is available on the website.

Figure 4.28(a) shows a precision half-wave rectifier circuit consisting of a diode placed in the negative-feedback path of an op amp, with R being the rectifier load resistance. The op amp, of course, needs power supplies, but for simplicity, these are not shown in the circuit diagram. The circuit works like this: If v_I goes positive, the output voltage v_A of the op amp will go positive and the diode will conduct, thus establishing a closed feedback path between the op amp's output terminal and the negative input terminal. This negative-feedback path will cause a virtual short circuit to appear between the two input terminals of the op amp. Thus the voltage at the negative input terminal, which is also the output voltage v_O, will equal (to within a few millivolts) that at the positive input terminal, which is the input voltage v_I,

$$v_O = v_I \quad v_I \geq 0$$

Note that the offset voltage ($\simeq 0.7$ V) exhibited in the simple half-wave rectifier circuit of Fig. 4.21 is no longer present. For the op-amp circuit to start operating, v_I has to exceed only a negligibly small voltage equal to the diode drop divided by the op amp's open-loop gain. In other words, the straight-line transfer characteristic v_O–v_I almost passes through the origin. This makes this circuit suitable for applications involving very small signals.

Now let's see what happens when v_I goes negative. The op amp's output voltage v_A will tend to follow and go negative. This will reverse-bias the diode, and no current will flow through resistance R, causing v_O to remain equal to 0 V. Thus, for $v_I < 0$, $v_O = 0$. Since in

[8]This section requires knowledge of operational amplifiers (Chapter 2).

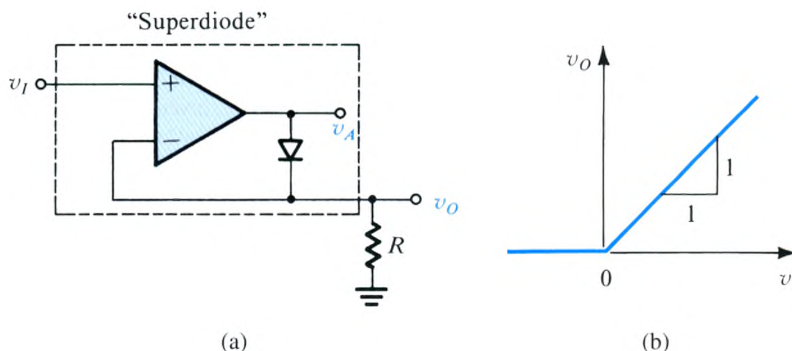


Figure 4.28 **(a)** The "superdiode" precision half-wave rectifier and **(b)** its almost-ideal transfer characteristic. Note that when $v_I > 0$ and the diode conducts, the op amp supplies the load current, and the source is conveniently buffered, an added advantage. Not shown are the op-amp power supplies.

this case the diode is off, the op amp will be operating in an open-loop fashion, and its output will be at its negative saturation level.

The transfer characteristic of this circuit will be that shown in Fig. 4.28(b), which is almost identical to the ideal characteristic of a half-wave rectifier. The nonideal diode characteristics have been almost completely masked by placing the diode in the negative-feedback path of an op amp. This is another dramatic application of negative feedback, a subject we will study formally in Chapter 11. The combination of diode and op amp, shown in the dashed box in Fig. 4.28(a), is appropriately referred to as a "superdiode."

EXERCISES

4.26 Consider the **operational rectifier** or superdiode circuit of Fig. 4.28(a), with $R = 1\ \text{k}\Omega$. For v_I = 10 mV, 1 V, and –1 V, what are the voltages that result at the rectifier output and at the output of the op amp? Assume that the op amp is ideal and that its output saturates at ±12 V. The diode has a 0.7-V drop at 1-mA current.

Ans. 10 mV, 0.59 V; 1 V, 1.7 V; 0 V, –12 V

4.27 If the diode in the circuit of Fig. 4.28(a) is reversed, find the transfer characteristic v_O as a function of v_I.

Ans. $v_O = 0$ for $v_I \geq 0$; $v_O = v_I$ for $v_I \leq 0$

A 4.7 Other Diode Applications

In this section we will consider additional nonlinear circuit applications of diodes along with special types of diodes worth noting.

4.7.1 The Clamped Capacitor and Bootstrapping

If in the basic peak-rectifier circuit we take the output across the diode rather than across the capacitor, we get an interesting circuit with important applications. The circuit, called a dc restorer, is shown in Fig. 4.29 fed with a square wave. Because of the polarity in which the diode is connected, the capacitor will charge to a voltage v_C with the polarity indicated in Fig. 4.29 and equal to the magnitude of the most negative peak of the input signal, neglecting the diode's voltage drop. Subsequently, the diode turns off and the capacitor retains its voltage indefinitely. If, for instance, the input square wave has the arbitrary levels –6 V and +4 V, then v_C will be equal to 6 V. Now, since the output voltage v_O is given by

$$v_O = v_I + v_C$$

it follows that the output waveform will be identical to that of the input, except that it is shifted upward by v_C volts. In our example the output will thus be a square wave with levels of 0 V and +10 V.

Another way of visualizing the operation of the circuit in Fig. 4.29 is to note that because the diode is connected across the output with the polarity shown, it prevents the output voltage from going below 0 V (by conducting and charging up the capacitor, thus causing the output to rise to 0 V), but this connection will not constrain the positive excursion of v_O. The output waveform will therefore have its lowest peak *clamped* to 0 V, which is why the circuit is called a **clamped capacitor**. It should be obvious that reversing the diode polarity will provide an output waveform whose highest peak is clamped to 0 V. In either case, the output waveform will have a finite average value or dc component that is entirely unrelated to the average value of the input waveform. As an application, consider a pulse signal being transmitted through a capacitively coupled or ac-coupled system. The capacitive coupling will cause the pulse train to lose whatever dc component it originally had. Feeding the resulting pulse waveform to a clamping circuit provides it with a well-determined dc component, a process known as **dc restoration**. This is why the circuit is also called a **dc restorer**.

Restoring dc is useful because the dc component or average value of a pulse waveform is an effective measure of its **duty cycle**.[9] The duty cycle of a pulse waveform can be modulated (in a process called **pulsewidth modulation**) and made to carry information. In such a system, detection or demodulation could be achieved simply by feeding the received

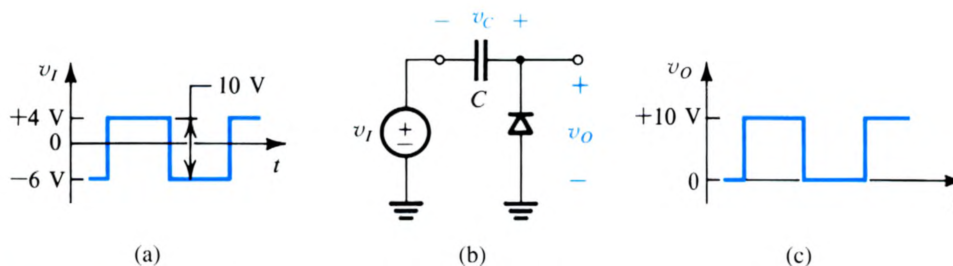


Figure 4.29 The clamped capacitor or dc restorer with a square-wave input and no load.

[9]The duty cycle of a pulse waveform is the proportion of each cycle occupied by the pulse. In other words, it is the pulse width expressed as a fraction of the pulse period.

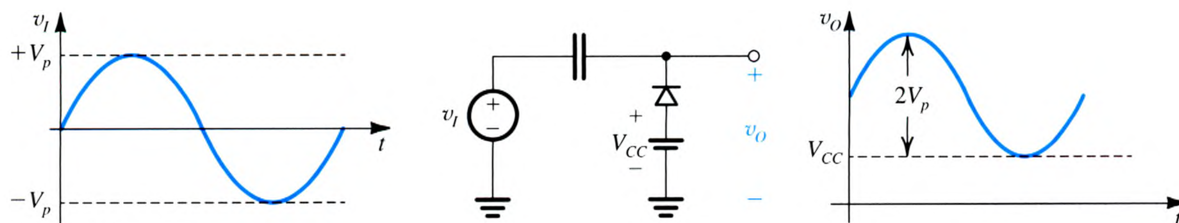
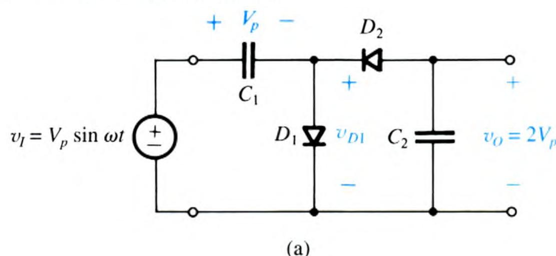
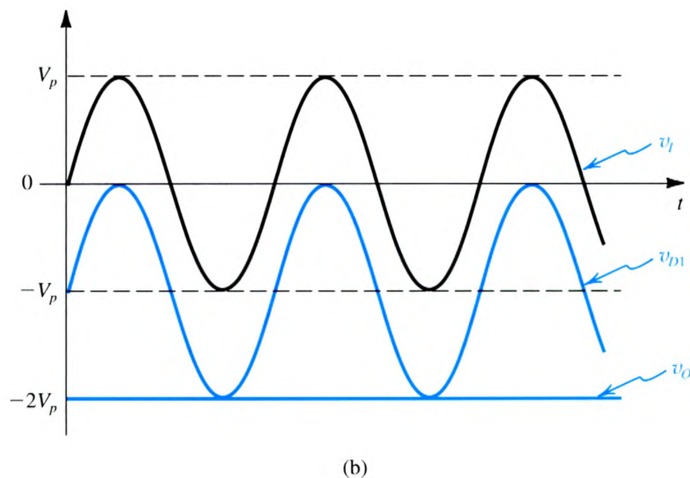


Figure 4.30 Simple bootstrapping circuit using a clamped capacitor.



(a)



(b)

Figure 4.31 Voltage doubler: **(a)** circuit; **(b)** waveforms of the input voltage, the voltage across D_1, and the output voltage $v_o = -2V_p$.

pulse waveform to a dc restorer and then using a simple *RC* low-pass filter to separate the average of the output waveform from the superimposed pulses.

The clamped capacitor may be modified so that the clamping voltage is V_{CC} instead of ground. Figure 4.30 is an example of a simple **bootstrapping circuit**, so named because the voltage v_I is "bootstrapped" onto V_{CC}. Bootstrapping is useful when a signal must be referenced to some potential other than ground, or when a voltage must track time variations in the voltage on another node.

4.7.2 The Voltage Doubler

Figure 4.31(a) shows a circuit made up of two sections in cascade: a clamped capacitor formed by C_1 and D_1, and a peak rectifier formed by D_2 and C_2. When excited by a sinusoid of amplitude V_p the clamping section provides the voltage waveform v_{D1} shown,

assuming ideal diodes, in Fig. 4.31(b). Note that while the positive peaks are clamped to 0 V, the negative peak reaches $-2V_p$. In response to this waveform, the peak-detector section provides a dc voltage across capacitor C_2 equal to the negative peak of v_{D1}, that is, $-2V_p$. Because the output voltage is double the input peak, the circuit is known as a voltage doubler. The technique can be extended to provide output dc voltages that are higher multiples of V_p.

EXERCISES

4.28 If the diode in the circuit of Fig. 4.29 is reversed, what will the dc component of v_O become?
Ans. –5 V

4.29 The bootstrapping circuit in Fig. 4.30 is used with $V_{CC} = 5$ V and $V_p = 3$ V. Assuming an ideal model for the diode, what is the voltage on the capacitor? What is the peak inverse voltage on the diode?
Ans. 8 V; 6 V

4.7.3 Varactors

In Chapter 3 we learned that reverse-biased *pn* junctions exhibit a charge-storage effect that is modeled with the depletion-layer or junction capacitance C_j. As Eq. (3.47) indicates, C_j is a function of the reverse-bias voltage V_R. This dependence turns out to be useful in a number of applications, such as the automatic tuning of radio receivers. Special diodes are therefore fabricated to be used as voltage-variable capacitors known as **varactors**. These devices are optimized to make the capacitance a strong function of voltage by arranging that the grading coefficient m is 3 or 4.

EXERCISE

4.30 Consider a varactor with a capacitance $C_{j0} = 100$ fF at zero bias, $V_0 = 3$ V, and $m = 3$. Using Eq. (3.47), estimate the capacitance at 1 V and 3 V reverse bias.
Ans. 42.2 fF; 12.5 fF

4.7.4 Photodiodes

If a reverse-biased *pn* junction is illuminated, photons impacting the junction cause covalent bonds to break. Electron-hole pairs are then generated in the depletion layer. The electric field in the depletion region sweeps the liberated electrons to the *n* side and the holes to the *p* side, giving rise to a reverse current across the junction. This **photocurrent**, i_P, is proportional to the intensity of incident light. Such a diode, called a photodiode, can be used to convert light signals into electrical signals.

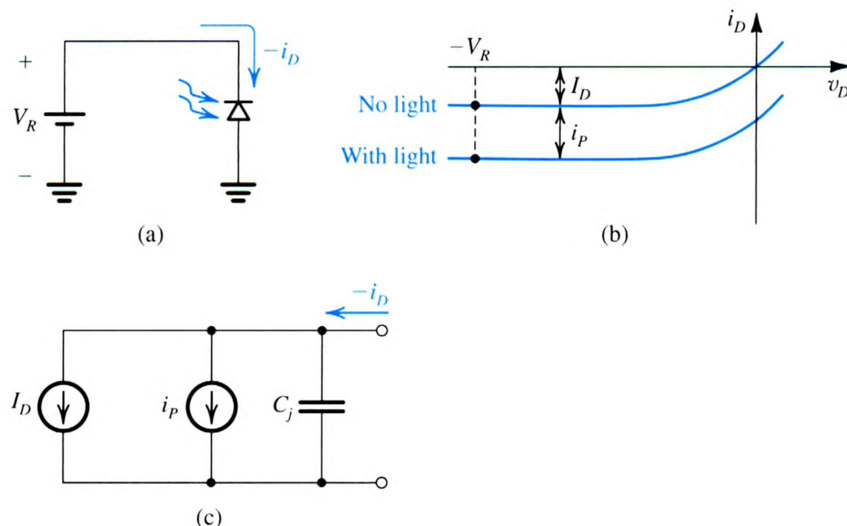


Figure 4.32 (**a**) Reverse-biased photodiode under illumination. (**b**) Shifting *i–v* characteristic of the photodiode under illumination. (**c**) Model of the photodiode under illumination with dark current I_D and photocurrent i_P.

The **responsivity**, *R*, of a photodiode is the reverse current produced per watt of incident light power, *P*. Hence,

$$i_P = R \cdot P \tag{4.33}$$

Higher responsivity is provided by photodiodes that are sensitive to longer wavelengths of light because each photon carries less energy, hence there are more photons (and more charge carriers generated) per watt of light. Responsivity is reduced if some of the photons are not absorbed (because either they are reflected or they pass through the photodiode) and if some of the charge carriers produced experience recombination before they can result in current at the photodiode terminals.

The symbol for a photodiode is shown in Fig. 4.32(a) with reverse bias, V_R. In the absence of light, a photodiode's *i–v* characteristic is that of a normal diode. The small reverse leakage current in a photodiode is referred to as **dark current**, I_D in Fig. 4.32(b), since it flows even in the absence of light. The *i–v* curve shifts toward more reverse current in proportion to the intensity of absorbed light. The model of a reverse-biased photodiode exposed to light in Fig. 4.32(c) shows dark current in parallel with the photocurrent. Thus, dark current places a lower limit on the level of light that may be readily detected. As with all diodes, the photodiode has an associated junction capacitance, C_j, which depends on the surface area of the photodiode and the applied reverse bias voltage, V_R.

Avalanche photodiodes are biased around their reverse breakdown voltage so that a single incident photon can set off an avalanche effect that results in more than one electron–hole pair. They therefore can exhibit much higher responsivities, but they require larger bias voltage.

The quantum of light carried by each photon depends upon its wavelength, and the amount of energy required to generate an electron–hole pair in a diode depends on the semiconductor material. Thus, different semiconductors absorb light at different wavelengths. Moreover, materials with a so-called indirect bandgap such as silicon are poor absorbers of light and therefore require a thicker depletion region to convert light into

current. Direct bandgap semiconductors absorb light within a very thin layer of material, making them better suited to high-speed operation. Thus photodiodes manufactured using a compound semiconductor[10] such as gallium arsenide are used when high-speed operation is required. For example, high-speed optical communication utilizes very narrow pulses of infrared light, only a few picoseconds in duration. The light is carried along glass fibers to the surface of a high-speed photodiode. When high-speed operation is not required, silicon photodiodes are often used to detect light in the visible spectrum because of their low cost.

Finally, note that without reverse bias, the illuminated photodiode functions as a solar cell. Usually fabricated from low-cost silicon, a solar cell converts light to electrical energy.

4.7.5 Light-Emitting Diodes (LEDs)

The light-emitting diode (LED) performs the opposite function of the photodiode: it converts a forward current into light. LEDs are very popular devices. They find application in the design of displays and for general lighting. They can be made to produce light in a variety of colors. Furthermore, LEDs can be designed to produce coherent light with high spectral purity. The resulting device is a laser diode. Laser diodes are used in optical communication systems and in optical data storage, among other applications.

You will recall from Chapter 3 that in a forward-biased *pn* junction, minority carriers are injected across the junction and diffuse into the p and n regions. The diffusing minority carriers then recombine with the majority carriers. In indirect bandgap semiconductors such as silicon, electrons must pass through an intermediate state as they move from the conduction band into a covalent bond (i.e., as they undergo recombination). In doing so, they transfer energy to the surrounding crystal lattice in the form of heat. However, in direct bandgap semiconductors, recombination can emit a photon. The energy, hence wavelength, of the photon depends upon the gap in energy between the two states, which is, in turn, related to the chemical composition of the semiconductor and the forward voltage drop required to sustain current in the *pn* junction. For example, red LEDs can be made from gallium arsenide phosphide and exhibit a forward voltage drop of approximately 1.8 V, whereas blue LEDs can be made from indium gallium nitride and have a forward voltage drop of 2.5–4.0 V.

The light emitted by an LED is proportional to the number of recombinations that take place, which in turn is proportional to the forward current in the diode. Commercial LEDs typically have an inherent series resistance that causes their i–v relationship to deviate from the exponential of a typical diode. Nevertheless, LED current increases rapidly with forward voltage, so when connecting them to voltage sources, an explicit series resistor is typically added.

[10]Whereas an elemental semiconductor, such as silicon, uses an element from column IV of the periodic table, a compound semiconductor uses a combination of elements from columns III and V or II and VI. For example, GaAs is formed of gallium (column III) and arsenic (column V) and is thus known as a III-V compound.

EXERCISES

4.31 A reverse biased photodiode with a responsivity of 0.5 A/W and 100 nA dark current is exposed to 1 mW of light. How much reverse current is produced? What if the incident light is 1 μW?
Ans. 0.5 mA; 0.6 μA

4.32 We wish to use a reverse-biased photodiode with responsivity of 0.3 A/W as a light sensor. Our objective is to sense light illuminating the photodiode with an intensity of 10 mW/m^2. What should the surface area of the photodiode be in order to produce 1 μA of current (neglecting dark current)? If at a particular reverse-bias voltage the photodiode exhibits 10 pF capacitance per mm^2 of surface area, estimate the photodiode capacitance.
Ans. 3.3 cm^2; 3.3 nF

4.33 We wish to operate three red LEDs in series as shown in Fig. E4.33 under a 9-V supply. They each have a rated forward voltage drop of 1.8 V at 20 mA. What value resistor, *R*, is required to sustain 20 mA?
Ans. 180 Ω

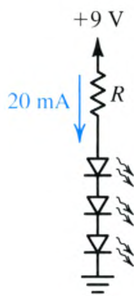


Figure E4.33

4.34 If, instead, green LEDs are used in Fig. E4.33, with a forward voltage drop of 2.2 V at 20 mA, what value resistor, *R*, is required?
Ans. 120 Ω

Combining an LED with a photodiode in the same package results in a device known as an optoisolator. The LED converts an electrical signal applied to the optoisolator into light, which the photodiode detects and converts back to an electrical signal at the output of the optoisolator. Use of the optoisolator provides complete electrical isolation between the electrical circuit that is connected to the isolator's input and the circuit that is connected to its output. Such isolation can be useful in reducing the effect of electrical interference on signal transmission within a system, making the optoisolator a frequently used device in the design of digital systems. It can also be used in the design of medical instruments to reduce the risk of electrical shock to patients.

Note that the optical coupling between an LED and a photodiode need not be accomplished inside a small package. Indeed, it can be implemented over a long distance using an optical fiber, as is done in fiber-optic communication links.

FROM INDICATION TO ILLUMINATION

Light-emitting diodes (LEDs) served only as low-powered status indicators until the early 1990s, when Isamu Akasaki, Hiroshi Amano, and Shuji Nakamura developed bright and efficient blue LEDs. Red and green LEDs had been around for a long time, but blue LEDs were needed to create white light. For their achievement, Akasaki, Amano, and Nakamura were awarded the Nobel Prize in 2014. Today, LEDs are lighting our way! LED bulbs are increasingly replacing both incandescent and fluorescent lighting everywhere. Incandescent bulbs are only 5% efficient in the conversion of electricity into light—the other 95% is dissipated as heat—while the light conversion efficiency of LEDs is 60%. Moreover, LEDs last 100 times longer (25,000 hours) than incandescent bulbs and 10 times longer than fluorescents. LEDs are thus conserving the earth's energy and material resources.

Summary

- In the forward direction, the ideal diode conducts any current forced by the external circuit while displaying a zero voltage drop. The ideal diode does not conduct in the reverse direction; any applied voltage appears as reverse bias across the diode.

- Diodes are often used to limit voltage excursions, for example, to protect sensitive circuits.

- The forward conduction of practical silicon-junction diodes is accurately characterized by the relationship $i = I_S e^{v/V_T}$.

- A silicon diode conducts a negligible current until the forward voltage is at least 0.5 V. Then the current increases rapidly, with the voltage drop increasing by 60 mV for every decade of current change.

- In many applications, a conducting diode is modeled as having a constant voltage drop of 0.7 V (for silicon).

- In the reverse direction, a silicon diode conducts a current on the order of 10^{-9} A. This current is much greater than I_S because of leakage effects, and it increases with the magnitude of reverse voltage.

- Beyond a certain value of reverse voltage (that depends on the diode), breakdown occurs, and current increases rapidly with a small corresponding increase in voltage. Zener diodes are designed to operate in the breakdown region.

- When operated with currents that deviate only slightly from a dc value I_D, a diode's exponential i–v relationship may be linearized. The resulting small-signal resistance is $r_d = V_T/I_D$. The approximation is accurate for variations in v_D of less than about 5 mV.

- We can use the nearly constant voltage drops produced by diodes, in both forward and reverse breakdown operation, to generate stable dc voltages. These are shunt voltage regulators.

- Rectifiers convert ac voltages into unipolar voltages. Half-wave rectifiers do this by passing the voltage in half of each cycle and blocking the opposite-polarity voltage in the other half of the cycle. Full-wave rectifiers accomplish the task by passing the voltage in half of each cycle and inverting the voltage in the other halfcycle.

- The bridge-rectifier circuit is the preferred full-wave rectifier configuration.

- The variation of the output waveform of the rectifier is reduced considerably by connecting a capacitor C across the output load resistance R. The resulting circuit is the peak rectifier. The output waveform then consists of a dc voltage almost equal to the peak of the input sine wave, V_p, on which is superimposed a ripple component of frequency $2f$ (in the full-wave case) and of peak-to-peak amplitude $V_r = V_p/2fCR$.

- Applying a time-varying waveform to a circuit consisting of a capacitor in series with a diode and taking the output across the diode provides a clamping function. Specifically, depending on the polarity of the diode, either the positive or negative peaks of the signal will be

clamped to the voltage at the other terminal of the diode (for example, ground). In this way the output waveform has a nonzero average or dc component, and the circuit is known as a dc restorer.

- By cascading a clamping circuit with a peak-rectifier circuit, we create a voltage doubler.
- Photodiodes are reverse-biased diodes that convert light into reverse current. Light-emitting diodes (LEDs) convert forward current into light. Depending on the color of light, LEDs are made from different semiconductors exhibiting different forward voltage drops greater than 0.7 V.

PROBLEMS

Computer Simulation Problems

SIM Problems identified by this icon are intended to demonstrate the value of using SPICE simulation to verify hand analysis and design, and to investigate important issues such as allowable signal swing and amplifier nonlinear distortion. Instructions to assist in setting up simulations for all the indicated problems can be found in the corresponding files on the website. Note that if a particular parameter value is not specified in the problem statement, you are to make a reasonable assumption.

Section 4.1: The Ideal Diode

4.1 An AA battery, whose Thévenin equivalent is a voltage source of 1.5 V and a resistance of 2 Ω, is connected to the terminals of an ideal diode. Describe two possible situations that result. What are the diode current and terminal voltage when (a) the connection is between the diode cathode and the positive terminal of the battery and (b) the anode and the positive terminal are connected?

4.2 For the circuits shown in Fig. P4.2 using ideal diodes, find the values of the voltages and currents indicated.

4.3 For the circuits shown in Fig. P4.3 using ideal diodes, find the values of the labeled voltages and currents.

4.4 In each of the ideal-diode circuits shown in Fig. P4.4, v_I is a 1-kHz, 5-V peak sine wave. Sketch the waveform resulting at v_O. What are its positive and negative peak values?

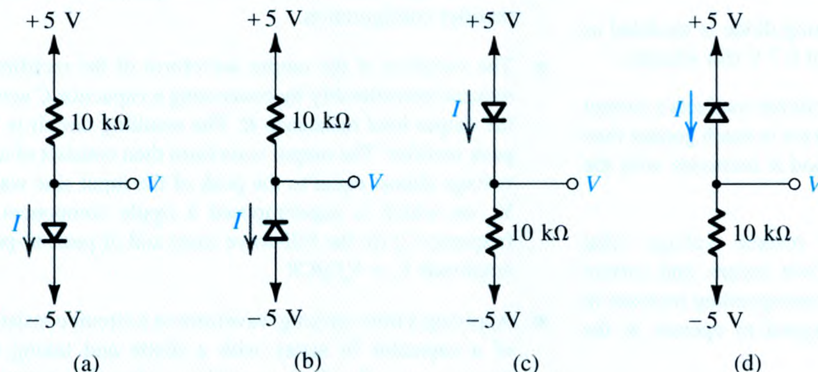


Figure P4.2

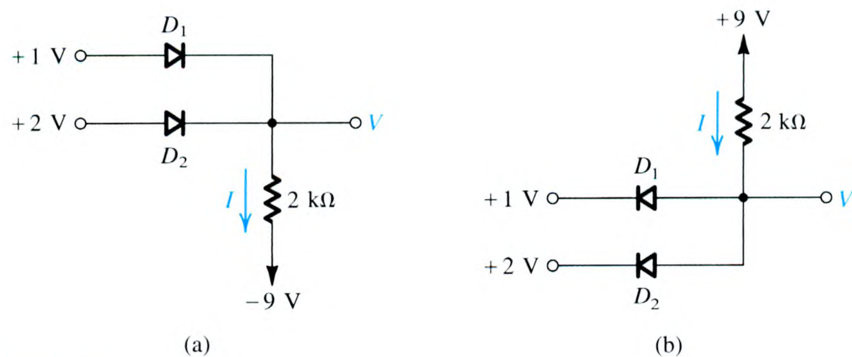


Figure P4.3

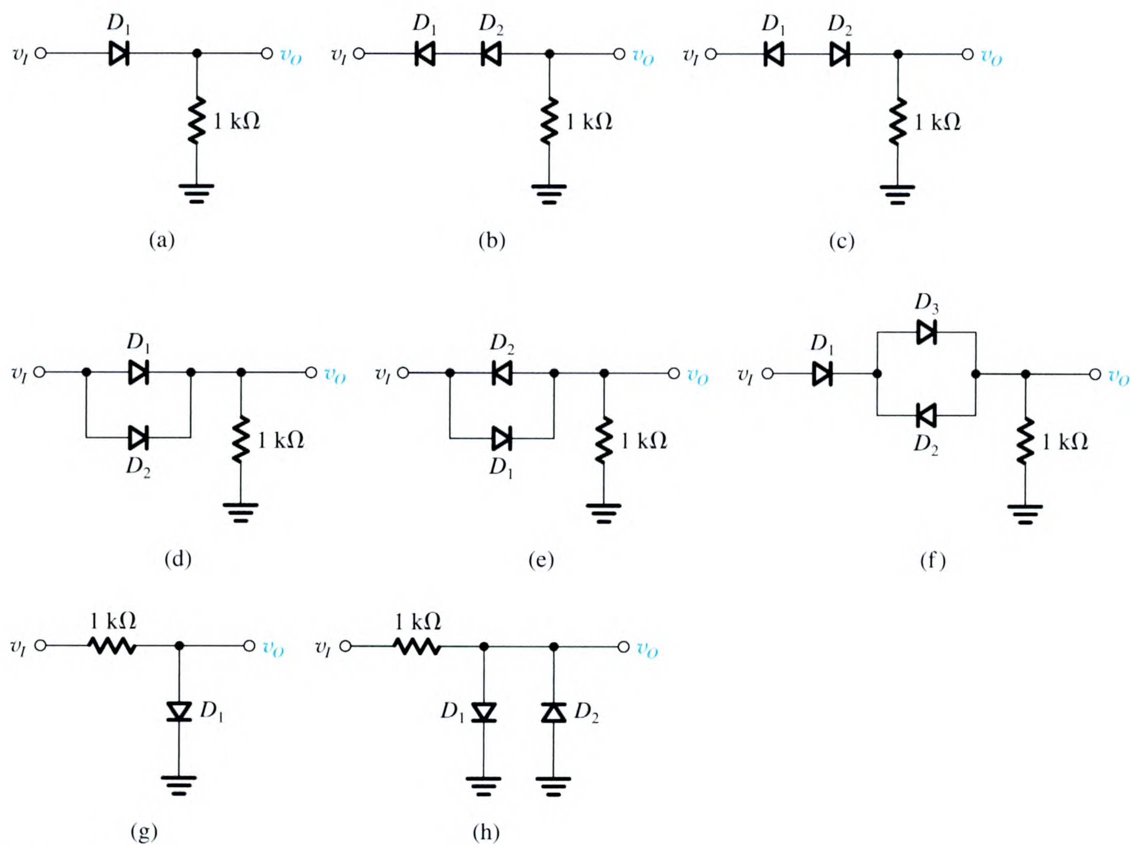


Figure P4.4

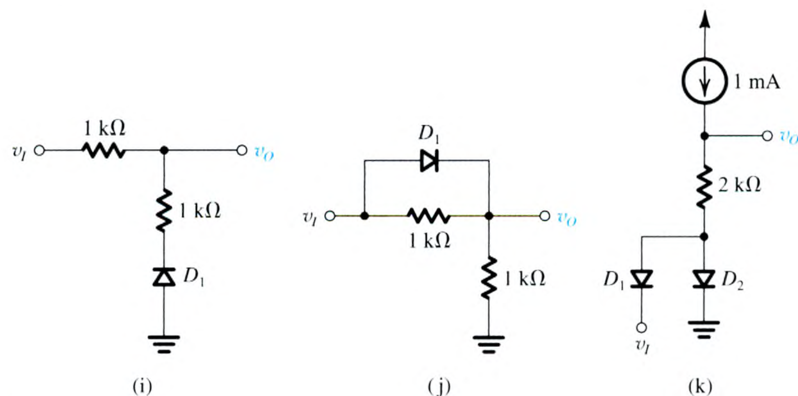


Figure P4.4 *continued*

4.5 The circuit shown in Fig. P4.5 is a model for a battery charger. Here v_I is a 10-V peak sine wave, D_1 and D_2 are ideal diodes, I is a 100-mA current source, and B is a 3-V battery. Sketch and label the waveform of the battery current i_B. What is its peak value? What is its average value? If the peak value of v_I is reduced by 10%, what do the peak and average values of i_B become?

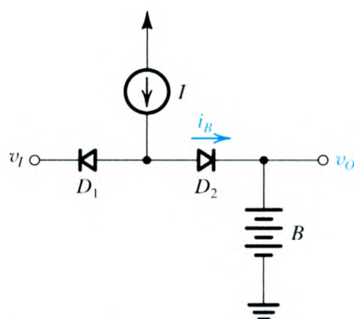


Figure P4.5

4.6 The circuits shown in Fig. P4.6 can function as logic gates for input voltages that are either high or low. Using "1" to denote the high value and "0" to denote the low value, prepare a table with four columns including all possible input combinations and the resulting values of X and Y. What logic function is X of A and B? What logic function is Y of A and B? For what values of A and B do X and Y have the same value? For what values of A and B do X and Y have opposite values?

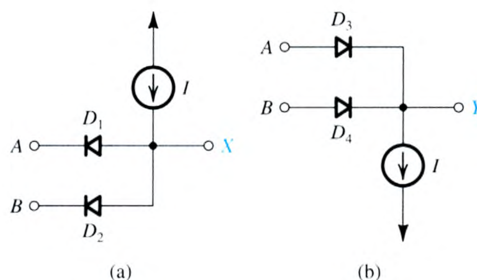


Figure P4.6



4.7 Assuming that the diodes in the circuits of Fig. P4.7 are ideal, find the values of the labeled voltages and currents.

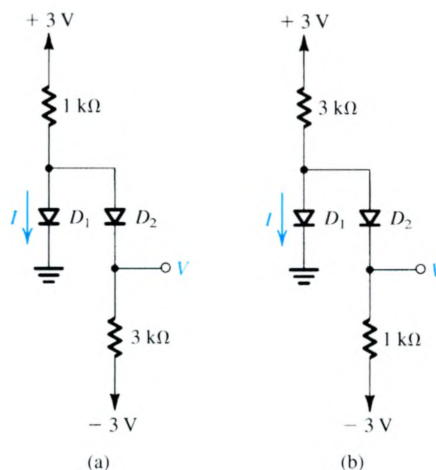


Figure P4.7

4.8 Assuming that the diodes in the circuits of Fig. P4.8 are ideal, utilize Thévenin's theorem to simplify the circuits and thus find the values of the labeled currents and voltages.

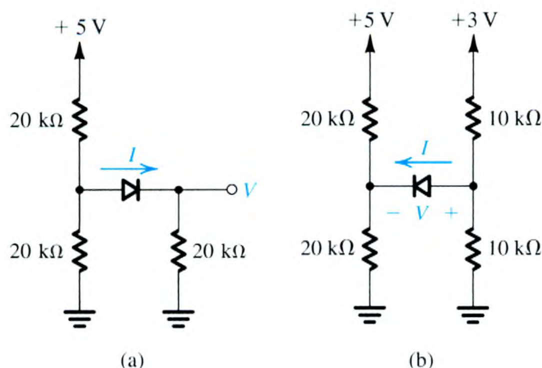


Figure P4.8

D 4.9 For the rectifier circuit of Fig. 4.3(a), let the input sine wave have 120-V rms value and assume the diode to be ideal. Select a suitable value for R so that the peak diode current does not exceed 40 mA. What is the greatest reverse voltage that will appear across the diode?

4.10 A symmetrical square wave of 10-V peak-to-peak amplitude and zero average is applied to a circuit resembling that in Fig. 4.3(a) and employing a 200-Ω resistor. What is the peak output voltage that results? What is the average output voltage that results? What is the peak diode current? What is the average diode current? What is the maximum reverse voltage across the diode?

4.11 We want to use the circuit of Fig. 4.3(a) with $R = 200\ \Omega$ and a square-wave input at v_I that alternates between +5 V and −3 V. Sketch the resulting waveform at v_O. What is the resulting peak and average current through the diode?

D 4.12 In the circuit of Fig. 4.4(a), select a new value for R so that the peak current through the diode is limited to 50 mA.

D *4.13 Design a battery-charging circuit, resembling that in Fig. 4.4(a) and using an ideal diode, in which current flows to the 12-V battery 25% of the time with an average value of 100 mA. What peak-to-peak sine-wave voltage is required? What resistance is required? What peak diode current flows? What peak reverse voltage does the diode endure? If resistors can be specified to only one significant digit, and the peak-to-peak voltage only to the nearest volt, what design would you choose to guarantee the required charging current? What fraction of the cycle does diode current flow? What is the average diode current? What is the peak diode current? What peak reverse voltage does the diode endure?

4.14 The circuit of Fig. P4.14 can be used in a signaling system using one wire plus a common ground return. At any moment, the input has one of three values: +3 V, 0 V, −3 V.

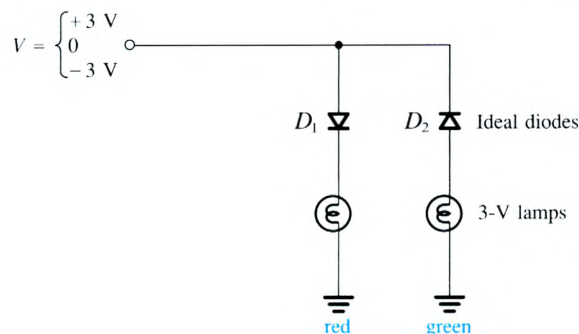


Figure P4.14

What is the status of the lamps for each input value? (Note that the lamps can be located apart from each other and that there may be several of each type of connection, all on one wire!)

4.15 The limiter circuit of Fig. 4.5(a) is connected with $V_+ = 3\ V$ and $R = 100\ \Omega$. The diode data sheet specifies a maximum reverse voltage of 10 V and maximum forward current of 50 mA. What are the maximum and minimum voltages that may be safely applied at v_I?

4.16 The limiter circuit shown in Fig. 4.5(a) has $V_+ = 5\ V$ and $R = 500\ \Omega$. Plot i_D versus the input voltage v_I. How does the plot change if a 1-kΩ load resistor is connected between v_O and ground?

4.17 The diodes in Fig. 4.6(a) must be able to withstand lightning strikes that cause the input voltage to temporarily spike to $v_I = \pm 200$ V. What peak current and peak reverse voltage must the diodes handle if $R = 100\ \Omega$?

Section 4.2: Terminal Characteristics of Junction Diodes

4.18 Calculate the value of the thermal voltage, V_T, at −55°C, 0°C, +40°C, and +125°C. At what temperature is V_T exactly 25 mV?

4.19 At what forward voltage does a diode conduct a current equal to $10^8 I_S$? In terms of I_S, what current flows in the same diode when its forward voltage is 0.7 V? Assume room temperature.

4.20 A diode for which the forward voltage drop is 0.7 V at 1.0 mA is operated at 0.6 V. What is the value of the current?

4.21 A particular diode is found to conduct 1 mA with a junction voltage of 0.7 V. What current will flow in this diode if the junction voltage is raised to 0.71 V? To 0.8 V? If the junction voltage is lowered to 0.69 V? To 0.6 V? What change in junction voltage will double the diode current?

4.22 The following measurements are taken on particular junction diodes for which V is the terminal voltage and I is the diode current. For each diode, estimate values of I_S and the terminal voltage at 10% of the measured current.

(a) $V = 0.750$ V at $I = 1.00$ A
(b) $V = 0.650$ V at $I = 1.00$ mA
(c) $V = 0.650$ V at $I = 10$ μA
(d) $V = 0.700$ V at $I = 10$ mA

4.23 Listed below are the results of measurements taken on several different junction diodes. For each diode, the data provided are the diode current I and the corresponding diode voltage V. In each case, estimate I_S and the diode current at 750 mV.

(a) 10.0 mA, 700 mV
(b) 1.0 mA, 700 mV
(c) 10 A, 800 mV
(d) 0.1 mA, 700 mV
(e) 10 μA, 600 mV

4.24 The circuit in Fig. P4.24 utilizes three identical diodes having $I_S = 10^{-14}$ A. Find the value of the current I required to obtain an output voltage $V_O = 2.0$ V at room temperature. If a current of 1 mA is drawn away from the output terminal by a load, what is the change in output voltage? What change in output voltage is caused by a 40°C increase in temperature?

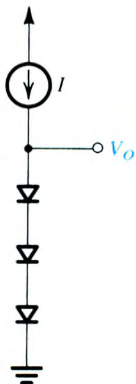


Figure P4.24

4.25 A junction diode is operated in a circuit in which it is supplied with a constant current I. What is the effect on the forward voltage of the diode if an identical diode is connected in parallel?

4.26 Two diodes with saturation currents I_{S1} and I_{S2} are connected in parallel with their cathodes joined together and connected to grounds. The two anodes are joined together and fed with a constant current I. Find the currents I_{D1} and I_{D2} that flow through the two diodes, and the voltage V_D that appears across their parallel combination.

VE 4.2 **4.27** In the circuit shown in Fig. P4.27, D_1 has 20 times the junction area of D_2. What value of V results? To obtain a value for V of 60 mV, what current I_2 is needed?

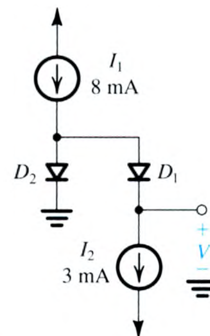


Figure P4.27

4.28 For the circuit shown in Fig. P4.28, the diodes are identical. Find the value of R for which $V = 50$ mV.

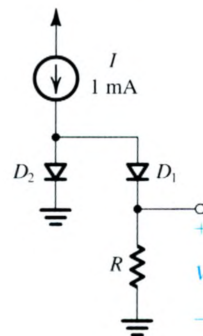


Figure P4.28

4.29 A diode fed with a constant current $I = 1$ mA has a voltage $V = 690$ mV at 20°C. Find the diode voltage at

problems with blue numbers are considered essential; * = difficult problem; ** = more difficult; *** = very challenging
SIM = simulation; **D** = design problem; ● = see related video example

$-20°C$ and at $+85°C$. If the current is increased to $I = 5$ mA, what is the impact on the diode voltage at $-20°C$, $20°C$, and $85°C$?

4.30 In the circuit shown in Fig. P4.30, D_1 is a large-area, high-current diode whose reverse leakage is high and independent of applied voltage, while D_2 is a much smaller, low-current diode. At an ambient temperature of 20°C, resistor R_1 is adjusted to make $V_{R1} = V_2 = 520$ mV. Subsequent measurement indicates that R_1 is 520 kΩ. What do you expect the voltages V_{R1} and V_2 to become at 0°C and at 60°C?

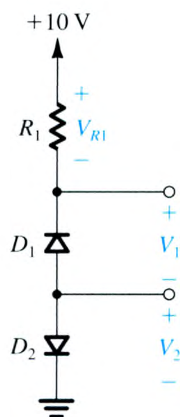


Figure P4.30

4.31 When a 10-A current is applied to a particular diode, it is found that the junction voltage immediately becomes 700 mV. However, as the power being dissipated in the diode raises its temperature, it is found that the voltage decreases and eventually reaches 600 mV. What is the apparent rise in junction temperature? What is the power dissipated in the diode in its final state? What is the temperature rise per watt of power dissipation? (This is called the thermal resistance.)

***4.32** A designer of an instrument that must operate over a wide supply-voltage range, noting that a diode's junction-voltage drop is relatively independent of junction current, considers the use of a large diode to establish a small relatively constant voltage. A power diode, for which the nominal current at 0.8 V is 10 A, is available. If the current source feeding the diode changes in the range 1 mA to 3 mA and if, in addition, the temperature changes by ±20°C, what is the expected range of diode voltage?

***4.33** As an alternative to the idea suggested in Problem 4.32, the designer considers a second approach to producing a relatively constant small voltage from a variable current supply: It relies on the ability to make quite accurate copies of any small current that is available (using a process called current mirroring). The designer proposes to use this idea to supply two diodes of different junction areas with equal currents and to measure their junction-voltage difference. Two types of diodes are available: for a forward voltage of 700 mV, one conducts 0.1 mA, while the other conducts 1 A. Now, for identical currents in the range of 1 mA to 3 mA supplied to each, what range of difference voltages result? What is the effect of a temperature change of ±20°C on this arrangement?

4.34 Find an expression for the voltage across a diode with a small series resistance, r, in terms of the current. If $r = 10$ Ω, what is the change in voltage when the current increases from 1 mA to 2 mA?

Section 4.3: Modeling the Diode

***4.35** Consider the graphical analysis of the diode circuit of Fig. 4.10 with $V_{DD} = 1$ V, $R = 1$ kΩ, and a diode having $I_S = 10^{-15}$ A. Calculate a small number of points on the diode characteristic in the vicinity of where you expect the load line to intersect it, and use a graphical process to refine your estimate of diode current. What value of diode current and voltage do you find? Analytically, find the voltage corresponding to your estimate of current. By how much does it differ from the graphically estimated value?

4.36 Use the iterative-analysis procedure to determine the diode current and voltage in the circuit of Fig. 4.10 for $V_{DD} = 1.5$ V, $R = 2$ kΩ, and a diode having $I_S = 10^{-15}$ A.

4.37 A "1-mA diode" (i.e., one that has $v_D = 0.7$ V at $i_D = 1$ mA) is connected in series with a 200-Ω resistor to a 1.0 V supply.

(a) Provide a rough estimate of the diode current you would expect.
(b) Estimate the diode current more closely using iterative analysis.

4.38 Consider the circuit in Fig. 4.10 with $V_{DD} = 3$ V and $R = 3$ kΩ.

(a) Find the current using a constant-voltage-drop model.
(b) What value of I_S is required to make this solution exact?
(c) Approximately how much will the current change from this value if I_S increases by a factor of 100?

D 4.39 A designer has a supply of diodes for which a current of 2 mA flows at 0.7 V. Using a 1-mA current source, the designer wishes to create a reference voltage of 1.3 V. Suggest a combination of series and parallel diodes that will do the job as well as possible. How many diodes are needed? What voltage is actually achieved?

4.40 Repeat Example 4.2 using the constant-voltage-drop ($V_D = 0.7$ V) diode model.

4.41 For the circuits shown in Fig. P4.2, using the constant-voltage-drop ($V_D = 0.7$ V) diode model, find the voltages and currents indicated.

4.42 For the circuits shown in Fig. P4.3, using the constant-voltage-drop ($V_D = 0.7$ V) diode model, find the voltages and currents indicated.


VE 4.3

4.43 For the circuits in Fig. P4.7, using the constant-voltage-drop ($V_D = 0.7$ V) diode model, find the values of the labeled currents and voltages.

4.44 For the circuits in Fig. P4.8, utilize Thévenin's theorem to simplify the circuits and find the values of the labeled currents and voltages. Assume that conducting diodes can be represented by the constant-voltage-drop model ($V_D = 0.7$ V).

D 4.45 Repeat Problem 4.9, representing the diode by the constant-voltage-drop ($V_D = 0.7$ V) model. How different is the resulting design?

4.46 Partial specifications of a collection of zener diodes are provided below. Complete the table by finding the missing value for each diode. Also, in each case, find the reverse voltage across the zener at 10 mA current.

	V_{ZT}	I_{ZT}	r_z	V_{Z0}	V_Z @ $I_Z = 10$ mA
(a)	3 V	2 mA	50 Ω		
(b)	5 V	5 mA		4.5 V	
(c)	6 V		150 Ω	5.7 V	
(d)		1 mA	100 Ω	9 V	

4.47 A zener diode is rated for 3-V operation at 5 mA with an incremental resistance of 100 Ω. It is connected to a voltage source V_S through a 1-kΩ resistor. What is the maximum voltage V_S that can be applied without exceeding the zener's maximum power rating of 40 mW? What is the voltage across the zener under this condition?

4.48 A constant current is passed through a 6-V zener diode having a TC of 1.5 mV/°C in series with a regular forward-conducting diode with a TC of −2 mV/°C. What is the resulting temperature coefficient of the voltage across both diodes?

Section 4.4: The Small-Signal Model

4.49 The small-signal model is said to be valid for voltage variations of about 5 mV. To what percentage current change does this correspond? (Consider both positive and negative signals.) What is the maximum allowable voltage signal (positive or negative) if the current change is to be limited to 10%?

4.50 In a particular circuit application, ten "20-mA diodes" (a 20-mA diode is a diode that provides a 0.7-V drop when the current through it is 20 mA) connected in parallel operate at a total current of 0.1 A. For the diodes closely matched, what current flows in each? What is the corresponding small-signal resistance of each diode and of the combination? Compare this with the incremental resistance of a single diode conducting 0.1 A. If each of the 20-mA diodes has a series resistance of 0.2 Ω associated with the wires bonded to the junction, what is the equivalent resistance of the 10 parallel-connected diodes? What connection resistance would a single diode need in order to be totally equivalent? (*Note:* This is why the parallel connection of real diodes can often be used to advantage.)

4.51 In the circuit shown in Fig. P4.51, I is a dc current and v_s is a sinusoidal signal. Capacitors C_1 and C_2 are very large; their function is to couple the signal to and from the diode but block the dc current from flowing into the signal source or the load (not shown). Use the diode small-signal model to show that the signal component of the output voltage is

$$v_o = v_s \frac{V_T}{V_T + IR_s}$$

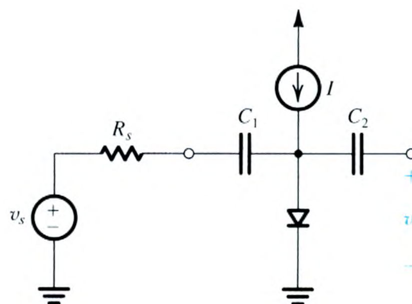


Figure P4.51

If $v_s = 10$ mV, find v_o for $I = 1$ mA, 0.1 mA, and 1 μA. Let $R_s = 1$ kΩ. At what value of I does v_o become one-half of v_s? Note that this circuit functions as a signal attenuator with the attenuation factor controlled by the value of the dc current I.

4.52 In the attenuator circuit of Fig. P4.51, let $R_s = 10$ kΩ. The diode is a 1-mA device; that is, it exhibits a voltage drop of 0.7 V at a dc current of 1 mA. For small input signals, what value of current I is needed for $v_o/v_s = 0.50$? 0.10? 0.01? 0.001? In each case, what is the largest input signal that can be used while ensuring that the signal component of the diode current is limited to ±10% of its dc current? What output signals correspond?

4.53 In the capacitor-coupled attenuator circuit shown in Fig. P4.53, I is a dc current that varies from 0 mA to 1 mA, and C_1 and C_2 are large coupling capacitors. For very small input signals, so that the diodes can be represented by their small-signal resistances r_{d1} and r_{d2}, give the small-signal equivalent circuit and thus show that $\frac{v_o}{v_i} = \frac{r_{d2}}{r_{d1} + r_{d2}}$ and hence that $\frac{v_o}{v_i} = I$, where I is in mA. Find v_o/v_i for $I =$ 0 μA, 1 μA, 10 μA, 100 μA, 500 μA, 600 μA, 900 μA, 990 μA, and 1 mA. Note that this is a signal attenuator whose transmission is linearly controlled by the dc current I.

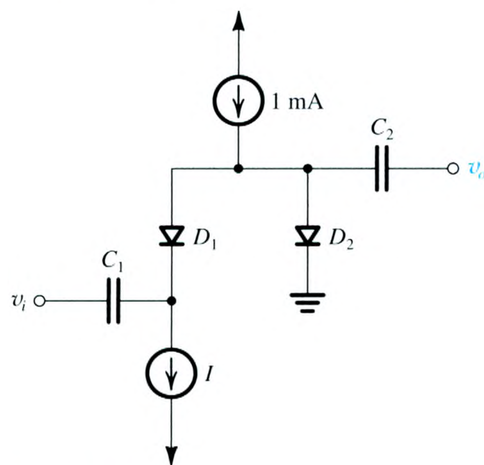


Figure P4.53

***4.54** In the circuit shown in Fig. P4.54, diodes D_1 through D_4 are identical, and each exhibits a voltage drop of 0.7 V at a 1-mA current.

(a) For small input signals (e.g., 10-mV peak), find the small-signal equivalent circuit and use it to determine values of the small-signal transmission v_o/v_i for various values of I: 0 μA, 1 μA, 10 μA, 100 μA, 1 mA, and 10 mA.

(b) For a forward-conducting diode, what is the largest signal-voltage magnitude that it can support while the corresponding signal current is limited to 10% of the dc bias current? Now, for the circuit in Fig. P4.54, for 10-mV peak input, what is the smallest value of I for which the diode currents remain within ±10% of their dc values?

(c) For $I = 1$ mA, what is the largest possible output signal for which the diode currents deviate by at most 10% of their dc values? What is the corresponding peak input? What is the total current in each diode?

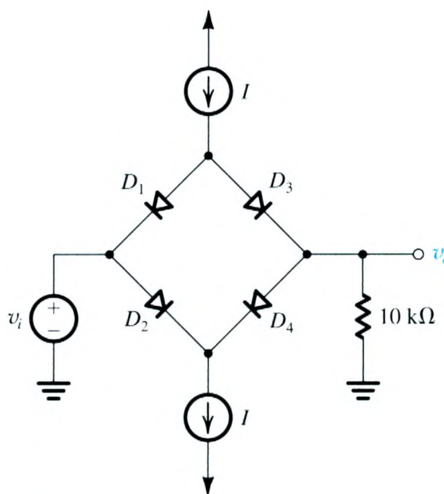


Figure P4.54

****4.55** In Problem 4.54 we investigated the operation of the circuit in Fig. P4.54 for small input signals. In this problem we wish to find the voltage-transfer characteristic (VTC) v_O versus v_I for $-12\text{ V} \le v_I \le 12\text{ V}$ for the case $I = 1$ mA and each of the diodes exhibits a voltage drop of 0.7 V at a current of 1 mA. Toward this end, use the diode exponential characteristic to construct a table that gives the values of: the current i_O in the 10-kΩ resistor, the current in each of the four diodes, the voltage drop across each of the four diodes, and the input voltage v_I, for $v_O = 0$, +1 V, +2 V, +5 V, +9 V, +9.9 V, +9.99 V, +10.5 V, +11 V, and +12 V. Use these data, with extrapolation to negative values of v_I and v_O, to sketch the required VTC. Also sketch the VTC that results if I is reduced to 0.5 mA. (*Hint*: From symmetry, observe that as v_O increases and i_O correspondingly increases, i_{D3} and i_{D2} increase by equal amounts and i_{D4} and i_{D1} decrease by [the same] equal amounts.)

SIM ***4.56** In the circuit shown in Fig. P4.56, I is a dc current and v_i is a sinusoidal signal with small amplitude (less than 10 mV) and a frequency of 100 kHz. Representing the diode by its small-signal resistance r_d, which is a function of I, sketch the small-signal equivalent circuit and use it to determine the sinusoidal output voltage V_o, and thus find the phase shift between V_i and V_o. Find the value of I that will provide a phase shift of –45°, and find the range of phase shift achieved as I is varied over the range of 0.1 times to 10 times this value.

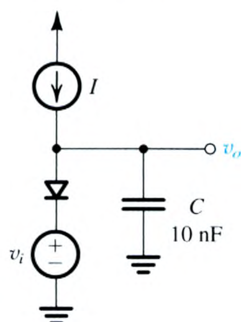


Figure P4.56

Section 4.5: Voltage Regulation

***4.57** Consider the voltage-regulator circuit shown in Fig. P4.57. The value of R is selected to obtain an output voltage V_O (across the diode) of 0.7 V.

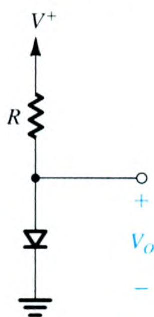


Figure P4.57

(a) Use the diode small-signal model to show that the change in output voltage corresponding to a change of 1 V in V^+ is

$$\frac{\Delta V_O}{\Delta V^+} = \frac{V_T}{V^+ + V_T - 0.7}$$

This quantity is known as the line regulation and is usually expressed in mV/V.

(b) Generalize the expression above for the case of m diodes connected in series and the value of R adjusted so that the voltage across each diode is 0.7 V (and $V_O = 0.7m$ V).

(c) Calculate the value of line regulation for the case $V^+ = 10$ V (nominally) and (i) $m = 1$ and (ii) $m = 3$.

***4.58** Consider the voltage-regulator circuit shown in Fig P4.57 under the condition that a load current I_L is drawn from the output terminal.

(a) If the value of I_L is sufficiently small that the corresponding change in regulator output voltage ΔV_O is small enough to justify using the diode small-signal model, show that

$$\frac{\Delta V_O}{I_L} = -(r_d \,\|\, R)$$

This quantity is known as the load regulation and is usually expressed in mV/mA.

(b) If the value of R is selected such that at no load the voltage across the diode is 0.7 V and the diode current is I_D, show that the expression derived in (a) becomes

$$\frac{\Delta V_O}{I_L} = \left(-\frac{V_T}{I_D}\right)\frac{V^+ - 0.7}{V^+ - 0.7 + V_T}$$

Select the lowest possible value for I_D that results in a load regulation whose magnitude is ≤ 5 mV/mA. If V^+ is nominally 10 V, what value of R is required? Also, specify the diode required in terms of its I_S.

(c) Generalize the expression derived in (b) for the case of m diodes connected in series and R adjusted to obtain $V_O = 0.7m$ V at no load.

D *4.59 Design a diode voltage regulator to supply 1.5 V to a 1.5-kΩ load. Use two diodes specified to have a 0.7-V drop at a current of 1 mA. The diodes are to be connected to a +5-V supply through a resistor R. Specify the value for R. What is the diode current with the load connected? Find and compare the power dissipated in the diodes to the power dissipated in the load. What is the increase resulting in the output voltage when the load is disconnected? What change results if the load resistance is reduced to 1 kΩ? To 750 Ω? To 500 Ω? (*Hint*: Use the small-signal diode model to calculate all changes in output voltage.)

D *4.60 A voltage regulator consisting of two diodes in series fed with a constant-current source is used as a replacement for a single carbon–zinc cell (battery) of nominal voltage 1.5 V. The regulator load current varies from 2 mA to 7 mA. Constant-current supplies of 5 mA, 10 mA, and 15 mA are available. Which would you choose, and why? What change in output voltage would result when the load current varies over its full range?

D 4.61 A designer requires a shunt regulator of approximately 20 V. Two kinds of zener diodes are available: 6.8-V devices with r_z of 10 Ω and 5.1-V devices with r_z of 25 Ω. For the two major choices possible, find the load regulation. In this calculation neglect the effect of the regulator resistance R.

4.62 A shunt regulator utilizing a zener diode with an incremental resistance of 10 Ω is fed through a 200-Ω resistor. If the raw supply changes by 1.0 V, what is the corresponding change in the regulated output voltage?

D 4.63 Design a 7.5-V zener regulator circuit using a 7.5-V zener specified at 10 mA. The zener has an incremental resistance $r_z = 30\ \Omega$ and a knee current of 0.5 mA. The regulator operates from a 10-V supply and delivers a nominal current of 5 mA to the load. What is the value of R you have chosen? What is the regulator output voltage when the supply is 10% high? Is 10% low? What is the output voltage when both the supply is 10% high and the load is removed? What is the largest load current that can be delivered while the zener operates at a current no lower than the knee current while the supply is 10% low? What is the load voltage in this case?

D *4.64 It is required to design a zener shunt regulator to provide a regulated voltage of about 10 V. The available 10-V, 1-W zener of type 1N4740 is specified to have a 10-V drop at a test current of 25 mA. At this current, its r_z is 7 Ω. The raw supply, V_S, available has a nominal value of 20 V but can vary by as much as ±25%. The regulator is required to supply a load current of 0 mA to 20 mA. Design for a minimum zener current of 5 mA.

(a) Find V_{Z0}.
(b) Calculate the required value of R.
(c) Find the line regulation. What is the change in V_O expressed as a percentage, corresponding to the ±25% change in V_S?
(d) Find the load regulation. By what percentage does V_O change from the no-load to the full-load condition?
(e) What is the maximum current that the zener in your design is required to conduct? What is the zener power dissipation under this condition?

Section 4.6: Rectifier Circuits

4.65 Consider the half-wave rectifier circuit of Fig. 4.21(a) with the diode reversed. Let v_S be a sinusoid with 5-V peak amplitude, and let $R = 2\ \text{k}\Omega$. Use the constant-voltage-drop diode model with $V_D = 0.7$ V.

(a) Sketch the transfer characteristic.
(b) Sketch the waveform of v_O.
(c) Find the average value of v_O.
(d) Find the peak current in the diode.
(e) Find the PIV of the diode.

4.66 Using the exponential diode characteristic, show that for v_S and v_O both greater than zero, the circuit of Fig. 4.21(a) has the transfer characteristic

$$v_O = v_S - v_D\,(\text{at } i_D = 1\text{ mA}) - V_T \ln(v_O/R)$$

where v_S and v_O are in volts and R is in kilohms. Note that this relationship can be used to obtain the voltage transfer characteristic v_O vs. v_S by finding v_S corresponding to various values of v_O.

SIM **4.67** Consider a half-wave rectifier circuit with a triangular-wave input of 6-V peak-to-peak amplitude and zero average, and with $R = 1\ \text{k}\Omega$. Assume that the diode can be represented by the constant-voltage-drop model with $V_D = 0.7$ V. Find the average value of v_O.

4.68 A half-wave rectifier circuit with a 500-Ω load operates from a 120-V (rms) 60-Hz household supply through a 12-to-1 step-down transformer. It uses a silicon diode that can be modeled to have a 0.7-V drop for any current. What is the peak voltage of the rectified output? For what fraction of the cycle does the diode conduct? What is the average output voltage? What is the average current in the load?


VE 4.4

4.69 A full-wave rectifier circuit with a 500-Ω load operates from a 120-V (rms) 60-Hz household supply through a 6-to-1 transformer having a center-tapped secondary winding. It uses two silicon diodes that can be modeled to have a 0.7-V drop for all currents. What is the peak voltage of the rectified output? For what fraction of a

cycle does each diode conduct? What is the average output voltage? What is the average current in the load?

4.70 A full-wave bridge-rectifier circuit with a 500-Ω load operates from a 120-V (rms) 60-Hz household supply through a 6-to-1 step-down transformer having a single secondary winding. It uses four diodes, each of which can be modeled to have a 0.7-V drop for any current. What is the peak value of the rectified voltage across the load? For what fraction of a cycle does each diode conduct? What is the average voltage across the load? What is the average current through the load?

D 4.71 It is required to design a full-wave rectifier circuit using the circuit of Fig. 4.22(a) to provide:

(a) an average output voltage of 10 V
(b) an average output voltage of 25 V
(c) a peak output voltage of 20 V

In each case find the required turns ratio of the transformer. Assume that a conducting diode has a voltage drop of 0.7 V. The ac line voltage is 120 V rms.

D 4.72 Repeat Problem 4.71 for the bridge-rectifier circuit of Fig. 4.23(a).

D 4.73 Consider the full-wave rectifier in Fig. 4.22(a) when the transformer turns ratio is such that the voltage across the entire secondary winding is 20 V rms. If the input ac line voltage is 220 V rms and it fluctuates by as much as ±10%, find the required PIV of the diodes. (Remember to use a factor of safety in your design.)

4.74 The circuit in Fig. P4.74 implements a complementary-output rectifier. Sketch and clearly label the waveforms of v_O^+ and v_O^-. Assume a 0.7-V drop across each conducting diode. If the magnitude of the average of each output is to be 12 V, find the required amplitude of the sine wave across the entire secondary winding. What is the PIV of each diode?

4.75 Augment the rectifier circuit of Problem 4.68 with a capacitor chosen to provide a peak-to-peak ripple voltage of (i) 10% of the peak output and (ii) 1% of the peak output. In each case:

(a) What average output voltage results?
(b) What fraction of the cycle does the diode conduct?
(c) What is the average diode current?
(d) What is the peak diode current?

4.76 Repeat Problem 4.75 for the rectifier in Problem 4.69.

4.77 Repeat Problem 4.75 for the rectifier in Problem 4.70.

D *4.78 It is required to use a peak rectifier to design a dc power supply that provides an average dc output voltage of 12 V on which a maximum of ±1-V ripple is allowed. The rectifier feeds a load of 200 Ω. The rectifier is fed from the line voltage (120 V rms, 60 Hz) through a transformer. The diodes available have 0.7-V drop when conducting. If the designer opts for the half-wave circuit:

(a) Specify the rms voltage that must appear across the transformer secondary.
(b) Find the required value of the filter capacitor.
(c) Find the maximum reverse voltage that will appear across the diode, and specify the PIV rating of the diode.

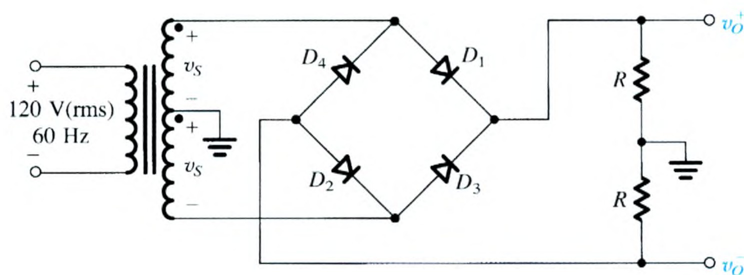


Figure P4.74

(d) Calculate the average current through the diode during conduction.
(e) Calculate the peak diode current.

D *4.79 Repeat Problem 4.78 for the case in which the designer opts for a full-wave circuit utilizing a center-tapped transformer.

D *4.80 Repeat Problem 4.78 for the case in which the designer opts for a full-wave bridge-rectifier circuit.

D *4.81 Consider a half-wave peak rectifier fed with a voltage v_S having a triangular waveform with 24-V peak-to-peak amplitude, zero average, and 1-kHz frequency. Assume that the diode has a 0.7-V drop when conducting. Let the load resistance $R = 100\ \Omega$ and the filter capacitor $C = 100\ \mu\text{F}$. Find the average dc output voltage, the time interval during which the diode conducts, and the maximum diode current.

D *4.82 Consider the circuit in Fig. P4.74 with two equal filter capacitors placed across the load resistors R. Assume that the diodes available exhibit a 0.7-V drop when conducting. Design the circuit to provide ±12-V dc output voltages with a peak-to-peak ripple no greater than 1 V. Each supply should be capable of providing 100-mA dc current to its load resistor R. Completely specify the capacitors, diodes, and the transformer.

4.83 The op amp in the precision rectifier circuit of Fig. P4.83 is ideal with output saturation levels of ±13 V. Assume that when conducting, the diode exhibits a constant voltage drop of 0.7 V. Find v_-, v_O, and v_A for:

(a) $v_I = +1$ V
(b) $v_I = +3$ V
(c) $v_I = -1$ V
(d) $v_I = -3$ V

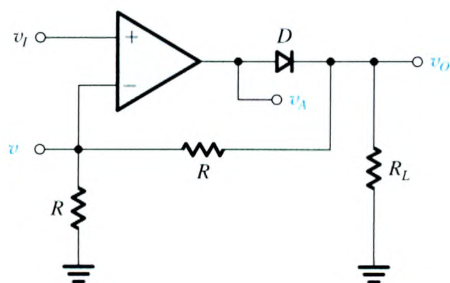


Figure P4.83

Also, plot the output waveform at v_O and find the average output voltage obtained when v_I is a symmetrical square wave of 1-kHz frequency, 5-V amplitude, and zero average.

4.84 The op amp in the circuit of Fig. P4.84 is ideal with output saturation levels of ±12 V. The diodes exhibit a constant 0.7-V drop when conducting. Find v_-, v_A, and v_O for:

(a) $v_I = +1$ V
(b) $v_I = +3$ V
(c) $v_I = -1$ V
(d) $v_I = -3$ V

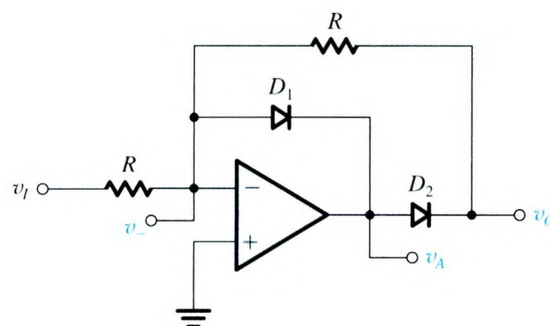


Figure P4.84

***4.85** The following amplitude-modulated signal is applied to the precision rectifier circuit in Fig. P4.85.

$$v_I = [1.5\ \text{V} + 1\ \text{V} \times \sin(2\pi 10^3 t)]\sin(2\pi 10^6 t)$$

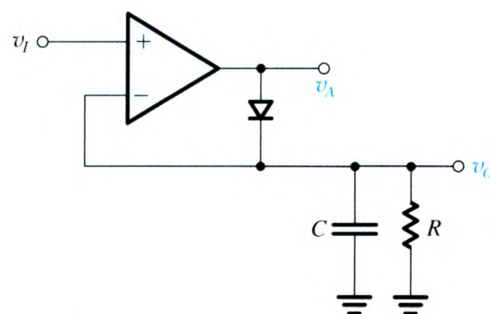


Figure P4.85

Sketch the waveform at the output, v_O, if the time constant CR is 0.1 ms. Repeat for $CR = 0.1\ \mu\text{s}$.

Section 4.7: Other Diode Applications

4.86 A clamped capacitor using an ideal diode with cathode grounded is supplied with a sine wave of 5-V rms. What is the average (dc) value of the resulting output?

4.87 The clamped capacitor in Fig. P4.87 is loaded by the resistor R. The input v_I is a symmetric 10-V peak-to-peak square wave at 100 kHz. Sketch the waveform at v_O if the time constant $CR = 10\ \mu\text{s}$.

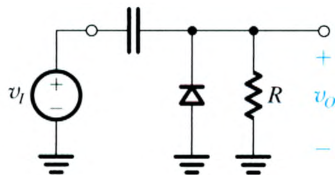


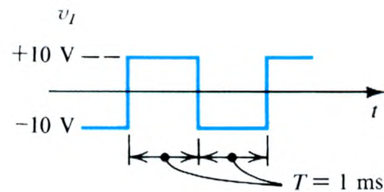
Figure P4.87

*4.88 For the circuits in Fig. P4.88, each utilizing an ideal diode (or diodes), sketch the output for the input shown. Label the most positive and most negative output levels. Assume $CR \gg T$.

4.89 A varactor has $C_{j0} = 150$ fF at zero bias, $V_0 = 3$ V, and $m = 3$. Find the range of bias voltages required to vary its capacitance from 80 fF to 20 fF.

4.90 A reverse-biased photodiode is specified to have a dark current of 100 pA and a responsivity of 0.5 A/W. It is connected to the transresistance amplifier shown in Fig. P4.90. Assume an ideal op amp.

(a) What is the reverse-bias voltage across the photodiode?
(b) What is the output voltage v_O with no illumination?
(c) What is the output voltage v_O with 10 μW of light incident on the photodiode?



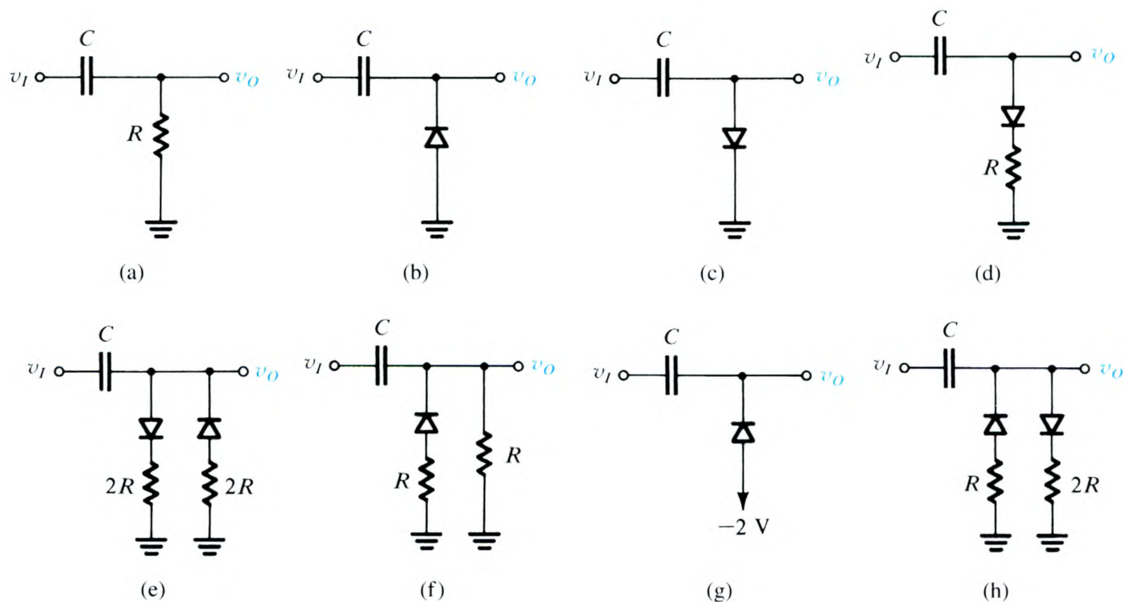


Figure P4.88

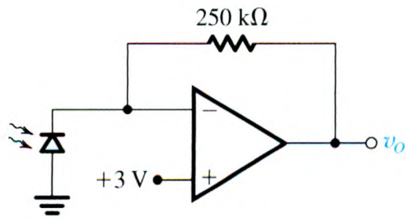


Figure P4.90

4.91 An LED is forward-biased by a 3-V battery through a resistor. Find the resistor value required to sustain

(a) 10 mA through a green LED at 2.2 V forward voltage drop.

(b) 10 mA through a red LED at 1.8 V forward voltage drop.

CHAPTER 5

MOS Field-Effect Transistors (MOSFETs)

IN THIS CHAPTER YOU WILL LEARN

- The physical structure of the MOS transistor and how it works.
- How the voltage between two terminals of the transistor controls the current that flows through the third terminal, and the equations that describe these current–voltage characteristics.
- How to analyze and design circuits that contain MOS transistors, resistors, and dc sources.
- Moore's law and how it has driven the advancement of microelectronics over the past 50 years.

Introduction

Having studied the junction diode, which is the most basic two-terminal semiconductor device, we turn our attention to three-terminal semiconductor devices. Three-terminal devices are far more useful than two-terminal ones because they can be used in a multitude of applications, from signal amplification to digital logic and memory. They operate by using the voltage between two terminals to control the current flowing in the third terminal. The three-terminal device can thus be used as a controlled source, which, as we learned in Chapter 1, is the basis for amplifier design. Also, in the extreme, we can use the control signal to cause the current in the third terminal to change from zero to a large value, thus allowing the device to act as a switch. As we shall see in Chapter 16, the switch is the basis for the logic inverter, the basic element of digital circuits.

There are two major types of three-terminal semiconductor devices:

- the metal-oxide-semiconductor field-effect transistor (MOSFET), which we cover in this chapter; and
- the bipolar junction transistor (BJT), which we study in Chapter 6.

Although both transistor types offer unique features and areas of application, the MOSFET has become by far the most widely used electronic device, especially in the design of integrated circuits (ICs), which are entire circuits fabricated on a single silicon chip.

Compared to BJTs, MOSFETs can be made quite small (i.e., requiring a small area on the silicon IC chip). They are simpler to manufacture (see Appendix A) and require less power to operate. Furthermore, circuit designers have found ingenious ways to implement digital and analog functions using MOSFETs almost exclusively (i.e., with very few or no resistors).

All of these properties have made it possible to pack large numbers of MOSFETs (many billion) on a single IC chip to create sophisticated, very-large-scale-integrated (VLSI) digital circuits such as those for memory and microprocessors. Analog circuits such as amplifiers and filters can also be implemented in MOS technology, albeit in smaller, less-dense chips. Also, both analog and digital functions are increasingly being implemented on the same IC chip, in what is known as mixed-signal design.

This chapter will help you become familiar with the MOSFET: its physical structure and operation, terminal characteristics, and dc circuit applications. This will give you a solid foundation for the application of the MOSFET in amplifier design (Chapter 7) and in digital circuit design (Chapter 16). Although discrete MOS transistors are available, and the material studied in this chapter will enable you to design discrete MOS circuits, our study of the MOSFET is strongly influenced by the fact that most of its applications are in integrated-circuit design. You will learn much more about the design of IC analog and digital MOS circuits throughout the rest of this book.

THE FIRST FIELD-EFFECT DEVICES

In 1925 a patent for a solid-state electric-field-controlled conductor was filed in Canada by Julius E. Lilienfeld, a physicist at the University of Leipzig, Germany. Other patent refinements followed in the United States in 1926 and 1928. Regrettably, no research papers were published. Consequently, in 1934 Oskar Heil, a German physicist working at the University of Cambridge, U.K., filed a patent on a similar idea. But all these early concepts of electric-field control of a semiconducting path languished because suitable technology was not available.

The invention of the bipolar transistor in 1947 at Bell Telephone Laboratories resulted in the speedy development of bipolar devices, a circumstance that further delayed the development of field-effect transistors. Although the field-effect device was described in a paper by William Shockley in 1952, it was not until 1960 that a patent on an insulated-gate field-effect device, the MOSFET, was filed by Dawon Kahng and Martin Atalla, also at Bell Labs. Clearly, the idea of field-effect control for amplification and switching has changed the world. With integrated-circuit chips today containing billions of MOS devices, MOS dominates the electronics world!

5.1 Device Structure and Physical Operation

The *enhancement-type* MOSFET is the most widely used field-effect transistor. Except for the last section, this chapter is devoted to the study of the enhancement-type MOSFET. We begin in this section by learning about its structure and physical operation. Then, in the next section, we will look at the current–voltage characteristics of the device.

5.1.1 Device Structure

Figure 5.1 shows the physical structure of the *n*-channel enhancement-type MOSFET. (The meaning of the names "enhancement" and "*n*-channel" will become apparent shortly.) The transistor is built on a *p*-type substrate, a single-crystal silicon wafer that provides physical support for the device (and for the entire circuit in the case of an integrated circuit). Two heavily doped *n*-type regions, indicated as the n^+ **source**[1] and the n^+ **drain** regions, are created in the substrate. A thin layer of silicon dioxide (SiO_2) of thickness t_{ox} (typically 1 nm to 10 nm),[2] which is an excellent electrical insulator, is grown on the surface of the substrate, covering the area between the source and drain regions. Metal is deposited on top of that oxide layer to form what's called the **gate electrode** of the device. Metal contacts are also made to the source region, the drain region, and the substrate, also known as the **body**.[3] This creates four terminals: the gate terminal (G), the source terminal (S), the drain terminal (D), and the substrate or body terminal (B).

At this point it should be clear that the name *metal-oxide-semiconductor FET* comes from its physical structure. The name, however, has become a general one and is used also for FETs that do not use metal for the gate electrode. Our description of MOSFET operation and characteristics applies regardless of which type of gate electrode is used.

Here we note that the gate electrode is electrically insulated from the device body (by the oxide layer). It is this insulation that causes the current in the gate terminal to be extremely small (of the order of 10^{-15} A).

---

[1]The notation n^+ indicates heavily doped *n*-type silicon. Conversely, n^- is used to denote lightly doped *n*-type silicon. Similar notation applies for *p*-type silicon.

[2]A nanometer (nm) is 10^{-9} m or 0.001 μm. A micrometer (μm), or micron, is 10^{-6} m.

[3]In Fig. 5.1, the contact to the body is shown on the bottom of the device. This will prove helpful in Section 5.4 when we look at a phenomenon known as the "body effect." It is important to note, however, that in actual ICs, contact to the body is made at a location on the top of the device.

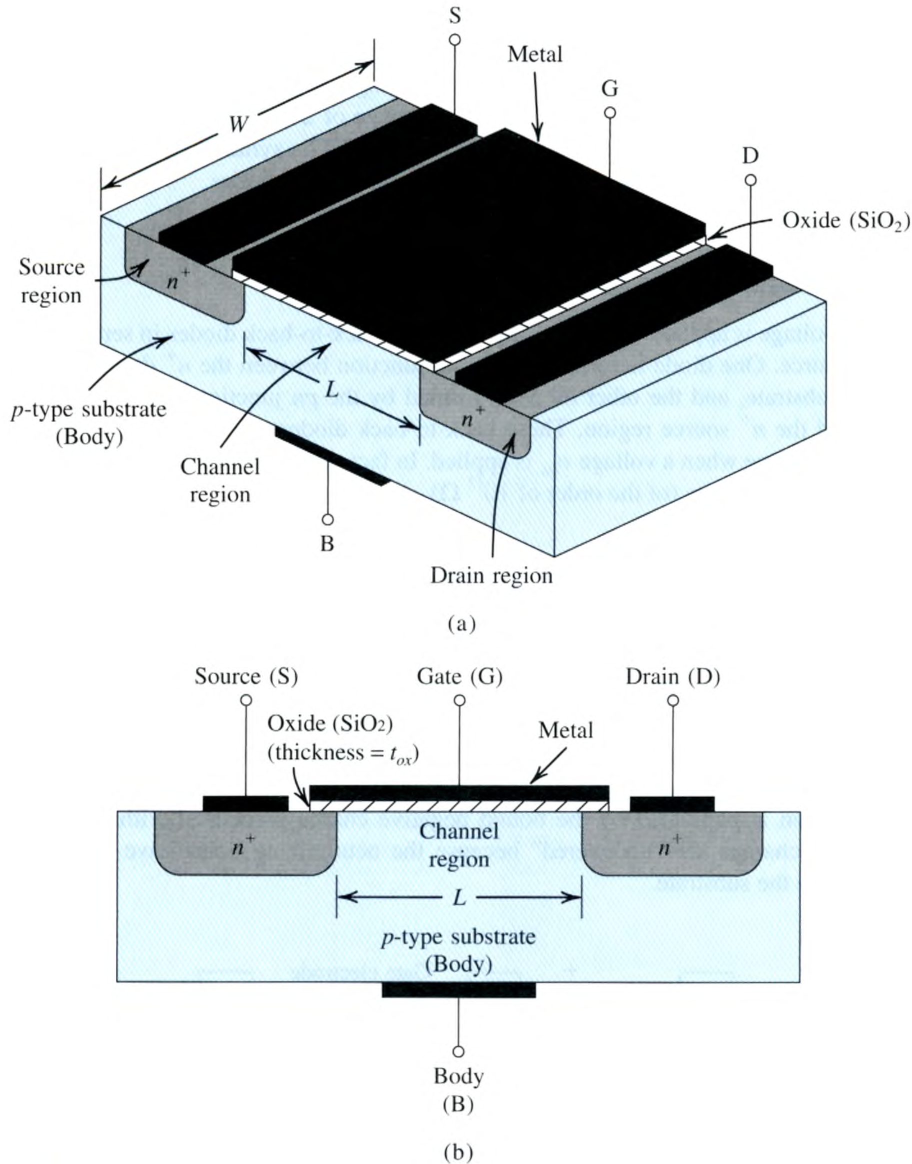


Figure 5.1 Physical structure of the n-channel enhancement-type MOSFET: **(a)** perspective view; **(b)** cross section. Typically, $L = 20$ nm to 1 μm, $W = 30$ nm to 100 μm, and the thickness of the oxide layer (t_{ox}) is in the range of 1 to 10 nm.

Notice that the substrate forms pn junctions with the source and drain regions. In normal operation these pn junctions are kept reverse biased at all times. Since, as we shall see shortly, the drain will always be at a positive voltage relative to the source, the two pn junctions can be effectively cut off by simply connecting the body terminal to the source terminal. We shall assume this to be the case in the following description of MOSFET operation. Thus, we will assume that the body has no effect on the device operation, and we can treat the MOSFET as a three-terminal device, with the terminals being the gate (G), the source (S), and the

drain (D). We will show that a voltage applied to the gate controls current flow between the source and the drain. This current will flow in the longitudinal direction from drain to source in the "channel region." Note that this region has a length L and a width W, two important parameters of the MOSFET. Typically, L is in the range of 20 nm to 1 μm, and W is in the range of 30 nm to 100 μm. Finally, note that the MOSFET is symmetrical; thus its source and drain can be interchanged with no change in device characteristics.

5.1.2 Operation with Zero Gate Voltage

When zero voltage is applied to the gate, we have two back-to-back diodes in series between drain and source. One diode is formed by the pn junction between the n^+ drain region and the p-type substrate, and the other diode is formed by the pn junction between the p-type substrate and the n^+ source region. These back-to-back diodes prevent current conduction from drain to source when a voltage v_{DS} is applied. In fact, the path between drain and source has a very high resistance (of the order of 10^{12} Ω).

5.1.3 Creating a Channel for Current Flow

Now let's consider the situation depicted in Fig. 5.2. Here we have grounded the source and the drain and applied a positive voltage to the gate. Since the source is grounded, the gate voltage appears in effect between gate and source and thus is denoted v_{GS}. The positive voltage on the gate causes, in the first instance, the free holes (which are positively charged) to be repelled from the area of the substrate just under the gate (the channel region). These holes are pushed downward into the substrate, leaving behind a carrier-depletion region. The depletion region is populated by the bound negative charge associated with the acceptor atoms. These charges are "uncovered" because the neutralizing holes have been pushed downward into the substrate.

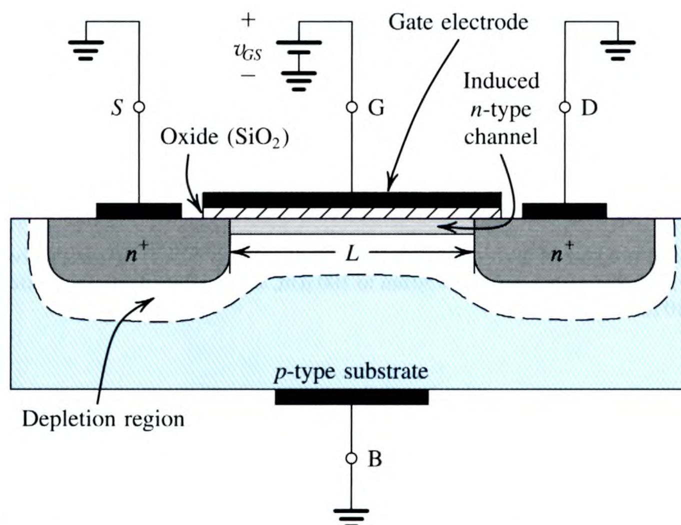


Figure 5.2 The enhancement-type NMOS transistor with a positive voltage applied to the gate. An n channel is induced at the top of the substrate beneath the gate.

As well, the positive gate voltage attracts electrons from the n^+ source and drain regions (where they are in abundance) into the channel region. When a sufficient number of electrons accumulate near the surface of the substrate under the gate, an n region is in effect created, connecting the source and drain regions to one another, as shown in Fig. 5.2. Now if we apply a voltage between drain and source, current flows through this induced n region, carried by the mobile electrons. The *induced n* region forms a **channel** for current flow from drain to source and is aptly called so. Correspondingly, the MOSFET of Fig. 5.2 is called an ***n*-channel MOSFET** or, alternatively, an **NMOS transistor**. Note that an n-channel MOSFET is formed in a p-type substrate: The channel is created by *inverting* the substrate surface from p type to n type.

The value of v_{GS} at which a sufficient number of mobile electrons accumulate under the gate to form a conducting channel is called the **threshold voltage**, V_t.[4] Obviously, V_t for an n-channel FET is positive. The value of V_t is controlled during device fabrication and typically lies in the range of 0.3 V to 1.0 V.

The gate and the channel region of the MOSFET form a parallel-plate capacitor, with the oxide layer acting as the capacitor dielectric. The positive gate voltage causes positive charge to accumulate on the top plate of the capacitor (the gate electrode). The corresponding negative charge on the bottom plate is formed by the electrons in the induced channel. An electric field thus develops in the vertical direction. It is this field that controls the amount of charge in the channel, and thus it determines the channel conductivity and, in turn, the current that flows through the channel when a voltage v_{DS} is applied. This is the origin of the name "field-effect transistor" (FET).

The voltage across this parallel-plate capacitor, that is, the voltage across the oxide, must exceed V_t for a channel to form. When $v_{DS} = 0$, as in Fig. 5.2, the voltage at every point along the channel is zero, and the voltage across the oxide (i.e., between the gate and the points along the channel) is uniform and equal to v_{GS}. The excess of v_{GS} over V_t is called the **effective voltage** or the **overdrive voltage** and is the quantity that determines the charge in the channel. In this book, we shall denote $(v_{GS} - V_t)$ by v_{OV},

$$v_{GS} - V_t \equiv v_{OV} \tag{5.1}$$

We can express the magnitude of the electron charge in the channel by

$$|Q| = C_{ox}(WL)v_{OV} \tag{5.2}$$

where C_{ox}, called the **oxide capacitance**, is the capacitance of the parallel-plate capacitor per unit gate area (in units of F/m^2), W is the width of the channel, and L is the length of the channel. The oxide capacitance C_{ox} is given by

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} \tag{5.3}$$

where ϵ_{ox} is the permittivity of the silicon dioxide,

$$\epsilon_{ox} = 3.9\epsilon_0 = 3.9 \times 8.854 \times 10^{-12} = 3.45 \times 10^{-11} \text{ F/m}$$

[4] Some texts use V_T to denote the threshold voltage. We use V_t to avoid confusion with the thermal voltage V_T.

The oxide thickness t_{ox} is determined by the process used to fabricate the MOSFET. As an example, for a process with $t_{ox} = 4$ nm,

$$C_{ox} = \frac{3.45 \times 10^{-11}}{4 \times 10^{-9}} = 8.6 \times 10^{-3} \text{ F/m}^2$$

It is much more convenient to express C_{ox} per micron squared. For our example, this yields 8.6 fF/μm^2, where fF denotes femtofarad (10^{-15} F). For a MOSFET fabricated in this technology with a channel length $L = 0.18\ \mu$m and width $W = 0.72\ \mu$m, the total capacitance between gate and channel is

$$C = C_{ox}WL = 8.6 \times 0.18 \times 0.72 = 1.1 \text{ fF}$$

Finally, note from Eq. (5.2) that as v_{OV} is increased, the magnitude of the channel charge increases proportionately. Sometimes this is depicted as an increase in the depth of the channel: the larger the overdrive voltage, the deeper the channel.

5.1.4 Applying a Small v_{DS}

Having induced a channel, we now apply a positive voltage v_{DS} between drain and source, as shown in Fig. 5.3. First consider the case where v_{DS} is small (i.e., 50 mV or so). The voltage v_{DS} causes a current i_D to flow through the induced n channel. Current is carried by free electrons traveling from *source* to *drain* (hence the names source and drain). By convention, the direction of current flow is opposite to that of the flow of negative charge. Thus the current in the channel, i_D, will be from drain to source, as shown in Fig. 5.3.

Now we need to calculate the value of i_D. Because v_{DS} is small, we can continue to assume that the voltage between the gate and various points along the channel remains approximately constant and equal to the value at the source end, v_{GS}. Thus, the effective voltage between the gate and the various points along the channel remains equal to v_{OV}, and the channel charge Q is still given by Eq. (5.2). To calculate the current i_D we first find the charge per unit channel length, which, from Eq. (5.2), is

$$\frac{|Q|}{L} = C_{ox}Wv_{OV} \tag{5.4}$$

The voltage v_{DS} establishes an electric field E across the length of the channel,

$$|E| = \frac{v_{DS}}{L} \tag{5.5}$$

This electric field in turn causes the channel electrons to drift toward the drain with a velocity given by

$$\text{Electron drift velocity} = \mu_n|E| = \mu_n\frac{v_{DS}}{L} \tag{5.6}$$

where μ_n is the mobility of the electrons at the surface of the channel, a physical parameter whose value depends on the fabrication process. We can now find the value of i_D by multiplying the charge per unit channel length (Eq. 5.4) by the electron drift velocity (Eq. 5.6),

$$i_D = \left[(\mu_n C_{ox})\left(\frac{W}{L}\right)v_{OV}\right]v_{DS} \tag{5.7}$$

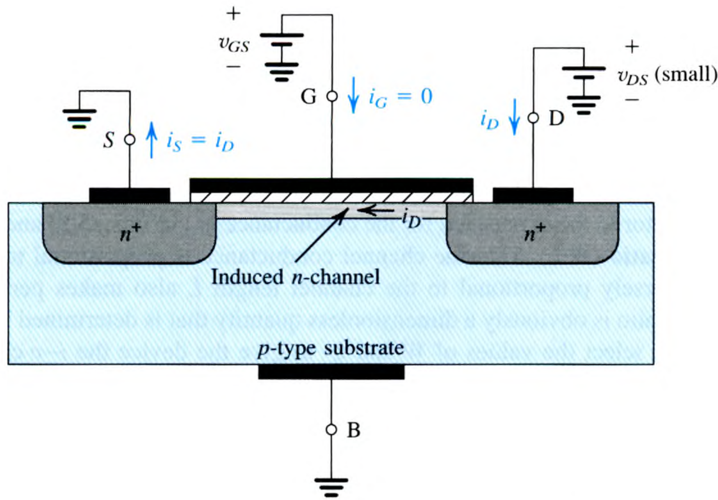


Figure 5.3 An NMOS transistor with $v_{GS} > V_t$ and with a small v_{DS} applied. The device acts as a resistance whose value is determined by v_{GS}. Specifically, the channel conductance is proportional to $v_{GS} - V_t$, and thus i_D is proportional to $(v_{GS} - V_t)v_{DS}$. Note that the depletion region is not shown (for simplicity).

Thus, for small v_{DS}, the channel behaves as a linear resistance whose value is controlled by the overdrive voltage v_{OV}, which in turn is determined by v_{GS}:

$$i_D = \left[(\mu_n C_{ox})\left(\frac{W}{L}\right)(v_{GS} - V_t)\right]v_{DS} \tag{5.8}$$

The conductance g_{DS} of the channel can be found from Eq. (5.7) or (5.8) as

$$g_{DS} = (\mu_n C_{ox})\left(\frac{W}{L}\right)v_{OV} \tag{5.9}$$

or

$$g_{DS} = (\mu_n C_{ox})\left(\frac{W}{L}\right)(v_{GS} - V_t) \tag{5.10}$$

Notice that the conductance is determined by the product of three factors: $(\mu_n C_{ox})$, (W/L), and v_{OV} (or equivalently, $v_{GS} - V_t$). To better understand MOSFET operation, we consider each of the three factors in turn.

The first factor, $(\mu_n C_{ox})$, is determined by the process technology used to fabricate the MOSFET. It is the product of the electron mobility, μ_n, and the oxide capacitance, C_{ox}. It makes physical sense for the channel conductance to be proportional to each of μ_n and C_{ox} and hence to their product, which is called the **process transconductance** parameter[5] and

[5]This name arises from the fact that $(\mu_n C_{ox})$ determines the transconductance of the MOSFET, as we will see in Chapter 7.

given the symbol k_n', where the subscript n denotes n channel,

$$k_n' = \mu_n C_{ox} \tag{5.11}$$

It can be shown that with μ_n having the dimensions of meters squared per volt-second ($m^2/V \cdot s$) and C_{ox} having the dimensions of farads per meter squared (F/m^2), the dimensions of k_n' are amperes per volt squared (A/V^2).

The second factor in the expression for the conductance g_{DS} in Eqs. (5.9) and (5.10) is the transistor **aspect ratio** (W/L). That the channel conductance is proportional to the channel width W and inversely proportional to the channel length L also makes perfect physical sense. The (W/L) ratio is obviously a dimensionless quantity that is determined by the device designer, who can select the values of W and L to give the device the i–v characteristics desired. For a given fabrication process, however, there is a minimum channel length, L_{min}. In fact, a given fabrication process is characterized by its smallest possible channel length, and the minimum is being continually reduced as technology advances. For instance, in 2019 the state-of-the-art in commercially available MOS technology was a 14-nm process, meaning that for this process the minimum channel length possible was 14 nm. Finally, we should note that the oxide thickness t_{ox} scales down with L_{min}. Thus, for a 0.13-μm technology, t_{ox} is 2.7 nm, but for the popular 65-nm technology, t_{ox} is about 2.2 nm. We will have more to say about technology scaling and Moore's law in Section 5.4.

The product of the process transconductance parameter k_n' and the transistor aspect ratio (W/L) is the **MOSFET transconductance parameter** k_n,

$$k_n = k_n'(W/L) = (\mu_n C_{ox})(W/L) \tag{5.12}$$

Just like k_n', the MOSFET transconductance parameter k_n has the dimensions of A/V^2.

The third term in the expression of the channel conductance g_{DS} is the overdrive voltage v_{OV}. This is hardly surprising, since v_{OV} directly determines the magnitude of electron charge in the channel. As we will see, v_{OV} is a very important circuit-design parameter. In this book, we will use v_{OV} and $v_{GS} - V_t$ interchangeably.

We conclude this subsection by noting that with v_{DS} kept small, the MOSFET behaves as a linear resistance r_{DS} whose value is controlled by the gate voltage v_{GS},

$$r_{DS} = \frac{1}{g_{DS}}$$

$$r_{DS} = \frac{1}{(\mu_n C_{ox})(W/L)v_{OV}} \tag{5.13a}$$

$$r_{DS} = \frac{1}{(\mu_n C_{ox})(W/L)(v_{GS} - V_t)} \tag{5.13b}$$

The operation of the MOSFET as a voltage-controlled resistance is illustrated in Fig. 5.4, which is a graph of i_D versus v_{DS} for various values of v_{GS}. Note that the resistance is infinite for $v_{GS} \le V_t$ and decreases as v_{GS} is increased above V_t. It is important to note that although v_{GS} is used as the parameter for the set of graphs in Fig. 5.4, the graphs in fact depend only on v_{OV} (and, of course, k_n).

By now you should understand that for the MOSFET to conduct, a channel has to be induced. Then, increasing v_{GS} above the threshold voltage V_t enhances the channel, hence the names **enhancement-mode operation** and **enhancement-type MOSFET**. Finally, we note that the current that leaves the source terminal (i_S) is equal to the current that enters the drain terminal (i_D), and the gate current $i_G = 0$.

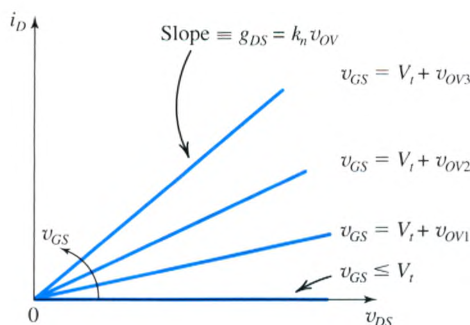


Figure 5.4 The i_D–v_{DS} characteristics of the MOSFET in Fig. 5.3 when the voltage applied between drain and source, v_{DS}, is kept small. The device operates as a linear resistance whose value is controlled by v_{GS}.

EXERCISE

5.1 A 0.18-μm fabrication process is specified to have $t_{ox} = 4$ nm, $\mu_n = 450$ cm^2/V · s, and $V_t = 0.5$ V. Find the value of the process transconductance parameter k_n'. For a MOSFET with minimum length fabricated in this process, find the required value of W so that the device exhibits a channel resistance r_{DS} of 1 kΩ at $v_{GS} = 1$ V.
Ans. 388 μA/V^2; 0.93 μm

5.1.5 Operation as v_{DS} Is Increased

Next let's consider the situation as v_{DS} is increased. For this purpose, let v_{GS} be held constant at a value greater than V_t; that is, let the MOSFET be operated at a constant overdrive voltage v_{OV}. In Fig. 5.5, you can see that v_{DS} appears as a voltage drop across the length of the channel. In other words, as we travel along the channel from source to drain, the voltage (measured relative to the source) increases from zero to v_{DS}. Thus the voltage between the gate and points along the channel decreases from $v_{GS} = V_t + v_{OV}$ at the source end to $v_{GD} = v_{GS} - v_{DS} = V_t + v_{OV} - v_{DS}$ at the drain end. Since the channel depth depends on this voltage, and specifically on the amount by which this voltage exceeds V_t, we find that the channel is no longer of uniform depth; rather, the channel takes the tapered shape shown in Fig. 5.5, being deepest at the source end (where the depth is proportional to v_{OV}) and shallowest at the drain end (where the depth is proportional to $v_{OV} - v_{DS}$). This point is further illustrated in Fig. 5.6.

As v_{DS} is increased, the channel becomes more tapered and its resistance increases correspondingly. Thus, the $i_D - v_{DS}$ curve does not continue as a straight line but bends, as shown in Fig. 5.7. The equation describing this portion of the $i_D - v_{DS}$ curve can be derived using the information in Fig. 5.6. Specifically, note that the charge in the tapered channel is proportional to the channel cross-sectional area shown in Fig. 5.6(b). This area in turn can be seen as proportional to $\frac{1}{2}[v_{OV}+(v_{OV}-v_{DS})]$ or $(v_{OV}-\frac{1}{2}v_{DS})$. Thus, the relationship between i_D and v_{DS} can be found by replacing v_{OV} in Eq. (5.7) by $(v_{OV}-\frac{1}{2}v_{DS})$,

$$i_D = k_n'\left(\frac{W}{L}\right)\left(v_{OV}-\frac{1}{2}v_{DS}\right)v_{DS} \tag{5.14}$$

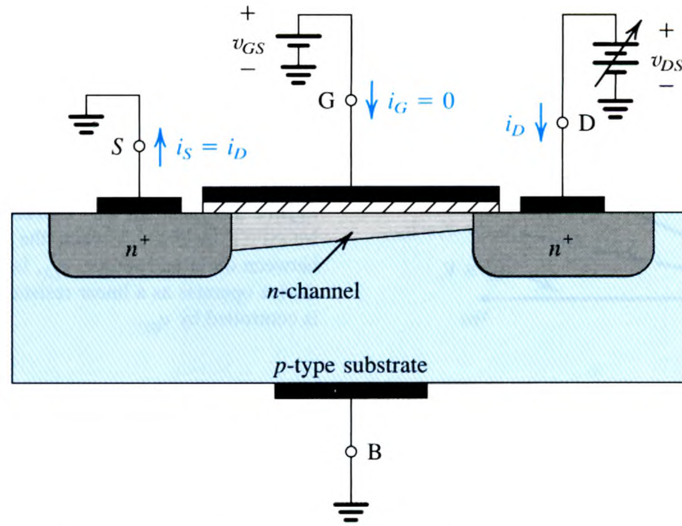


Figure 5.5 Operation of the MOSFET as v_{DS} is increased. The induced channel acquires a tapered shape, and its resistance increases as v_{DS} is increased. Here, v_{GS} is kept constant at a value > V_t; $v_{GS} = V_t + v_{OV}$. It is the widening of the depletion region (not shown) as a result of the increased v_{DS} that makes the channel shallower near the drain.

This relationship describes the semiparabolic portion of the $i_D - v_{DS}$ curve in Fig. 5.7. It applies to the entire segment down to $v_{DS} = 0$. Specifically, note that as v_{DS} is reduced, we can neglect $\frac{1}{2}v_{DS}$ relative to v_{OV} in the factor in parentheses, and the expression reduces to that in Eq. (5.7). The latter of course is an approximation and applies only for small v_{DS} (i.e., near the origin).

There is another useful interpretation of the expression in Eq. (5.14). From Fig. 5.6(a) we see that the average voltage along the channel is $\frac{1}{2}v_{DS}$. Thus, the average voltage that gives rise to channel charge and hence to i_D is no longer v_{OV} but $\left(v_{OV} - \frac{1}{2}v_{DS}\right)$, which is indeed the factor that appears in Eq. (5.14). Finally, we should note that Eq. (5.14) is frequently written in the alternate form

$$i_D = k_n'\left(\frac{W}{L}\right)\left(v_{OV}v_{DS} - \frac{1}{2}v_{DS}^2\right) \tag{5.15}$$

Also, we can replace v_{OV} by $(v_{GS} - V_t)$ and rewrite Eq. (5.15) as

$$i_D = k_n'\left(\frac{W}{L}\right)\left[(v_{GS} - V_t)v_{DS} - \frac{1}{2}v_{DS}^2\right] \tag{5.16}$$

5.1.6 Operation for $v_{DS} \geq v_{OV}$: Channel Pinch-Off and Current Saturation

Above we assumed that even though the channel became tapered, it still had a finite (nonzero) depth at the drain end. This is achieved by keeping v_{DS} sufficiently small that the voltage between the gate and the drain, v_{GD}, exceeds V_t. This is indeed the situation shown in Fig. 5.6(a). Note that for this situation to occur, v_{DS} must not exceed v_{OV}, for as $v_{DS} = v_{OV}$, $v_{GD} = V_t$, and the channel depth at the drain end reduces to zero.

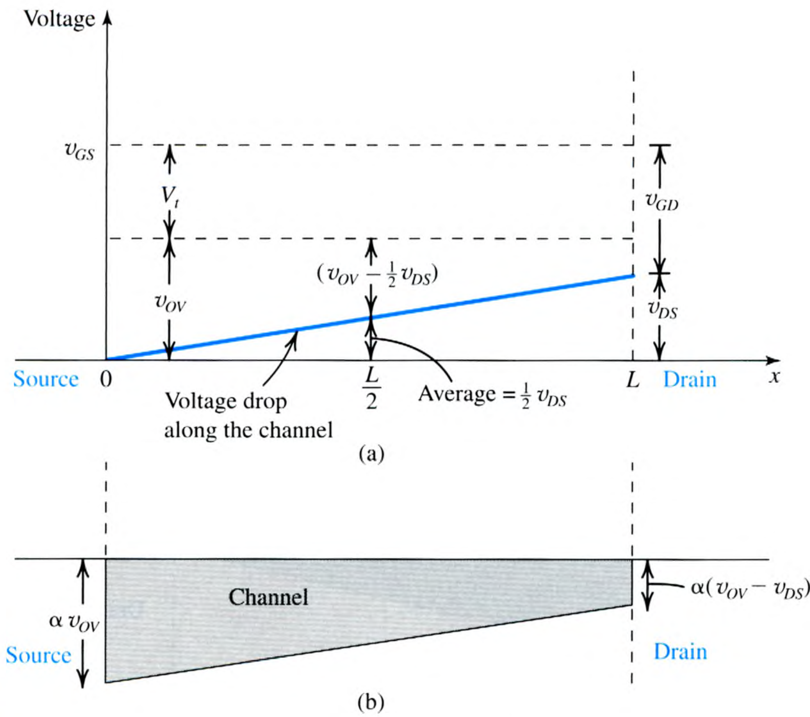


Figure 5.6 **(a)** For a MOSFET with $v_{GS} = V_t + v_{OV}$, applying of v_{DS} causes the voltage drop along the channel to vary linearly, with an average value of $\frac{1}{2}v_{DS}$ at the midpoint. Since $v_{GD} > V_t$, the channel still exists at the drain end. **(b)** The channel shape corresponding to the situation in **(a)**. While the depth of the channel at the source end is still proportional to v_{OV}, that at the drain end is proportional to $(v_{OV} - v_{DS})$.

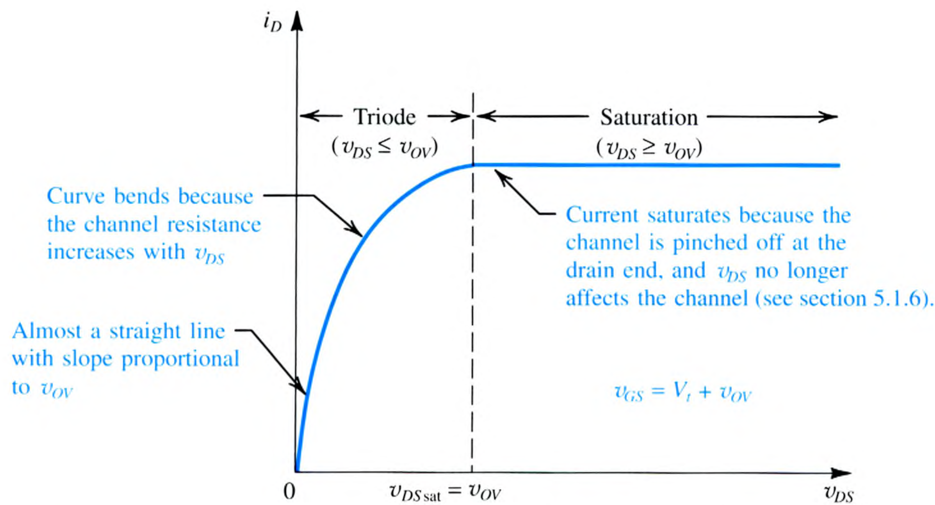


Figure 5.7 The drain current i_D versus the drain-to-source voltage v_{DS} for an enhancement-type NMOS transistor operated with $v_{GS} = V_t + v_{OV}$.

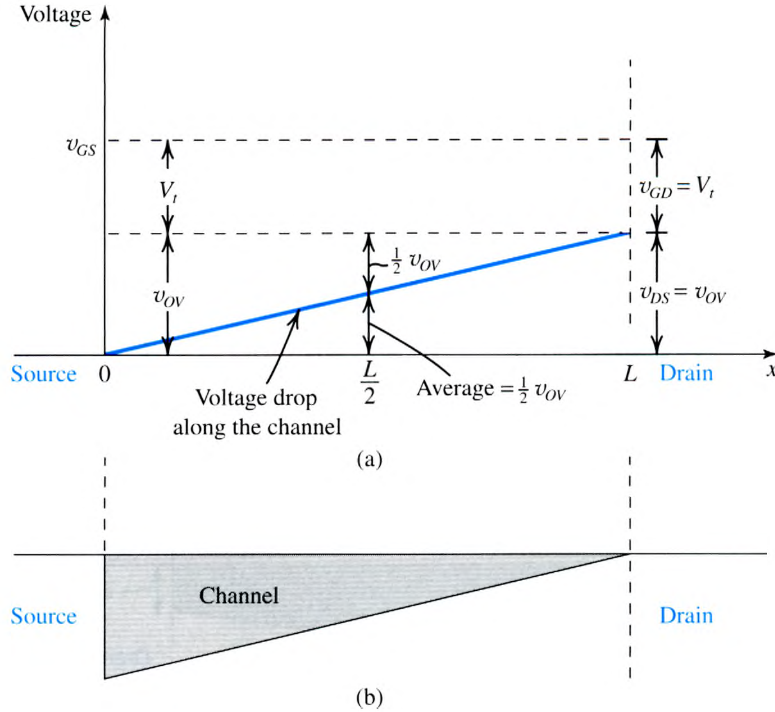


Figure 5.8 Operation of MOSFET with $v_{GS} = V_t + v_{OV}$, as v_{DS} is increased to v_{OV}. At the drain end, v_{GD} decreases to V_t and the channel depth at the drain end reduces to zero (pinch-off). At this point, the MOSFET enters the saturation mode of operation. Further increasing v_{DS} (beyond $v_{DSsat} = v_{OV}$) has no effect on the channel shape and i_D remains constant.

Figure 5.8 shows v_{DS} reaching v_{OV} and v_{GD} correspondingly reaching V_t. The zero depth of the channel at the drain end gives rise to the term **channel pinch-off**. Increasing v_{DS} beyond this value (i.e., $v_{DS} > v_{OV}$) has no effect on the channel shape and charge, and the current through the channel remains constant at the value reached for $v_{DS} = v_{OV}$. We say that the drain current **saturates** at the value found by substituting $v_{DS} = v_{OV}$ in Eq. (5.14),

$$i_D = \frac{1}{2} k_n' \left(\frac{W}{L} \right) v_{OV}^2 \tag{5.17}$$

The MOSFET is then said to have entered the **saturation region** (or, equivalently, the saturation mode of operation). The voltage v_{DS} at which saturation occurs is denoted v_{DSsat},

$$v_{DSsat} = v_{OV} = v_{GS} - V_t \tag{5.18}$$

Note that channel pinch-off does *not* mean channel blockage: Current continues to flow through the pinched-off channel, and the electrons that reach the drain end of the channel are accelerated through the depletion region that exists there (not shown in Fig. 5.5) and into the drain terminal. Any increase in v_{DS} above v_{DSsat} appears as a voltage drop across the depletion region. Thus, both the current through the channel and the voltage drop across it remain constant in saturation.

The saturation portion of the $i_D - v_{DS}$ curve is, as we might expect, a horizontal straight line, as shown in Fig. 5.7. Also indicated in Fig. 5.7 is the name of the region of operation

obtained with a continuous (non-pinched-off) channel, the **triode region**. This name is a carryover from the days of vacuum-tube devices, whose operation a FET resembles.

Finally, we can replace v_{OV} in Eq. (5.17) by $(v_{GS}-V_t)$ to obtain the alternate expression for saturation-mode i_D,

$$i_D = \frac{1}{2}k_n'\left(\frac{W}{L}\right)(v_{GS} - V_t)^2 \tag{5.19}$$

Example 5.1

Consider a process technology for which $L_{min} = 0.18\ \mu\text{m}$, $t_{ox} = 4\ \text{nm}$, $\mu_n = 450\ \text{cm}^2/\text{V}\cdot\text{s}$, and $V_t = 0.5\ \text{V}$.

(a) Find C_{ox} and k_n'.
(b) For a MOSFET with $W/L = 1.8\ \mu\text{m}/0.18\ \mu\text{m}$, calculate the values of v_{OV}, v_{GS}, and $v_{DS\text{min}}$ needed to operate the transistor in the saturation region with a current $i_D = 100\ \mu\text{A}$.
(c) For the device in (b), find the values of v_{OV} and v_{GS} required to cause the device to operate as a 1000-Ω resistor for very small v_{DS}.

Solution

(a)

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} = \frac{3.45 \times 10^{-11}}{4 \times 10^{-9}} = 8.63 \times 10^{-3}\ \text{F/m}^2$$
$$= 8.63\ \text{fF}/\mu\text{m}^2$$
$$k_n' = \mu_n C_{ox} = 450\ (\text{cm}^2/\text{V}\cdot\text{s}) \times 8.63\ (\text{fF}/\mu\text{m}^2)$$
$$= 450 \times 10^8\ (\mu\text{m}^2/\text{V}\cdot\text{s}) \times 8.63 \times 10^{-15}\ (\text{F}/\mu\text{m}^2)$$
$$= 388 \times 10^{-6}\ (\text{F/V}\cdot\text{s})$$
$$= 388\ \mu\text{A/V}^2$$

(b) For operation in the saturation region,

$$i_D = \frac{1}{2}k_n'\frac{W}{L}v_{OV}^2$$

Thus,

$$100 = \frac{1}{2} \times 388 \times \frac{1.8}{0.18}v_{OV}^2$$

which gives

$$v_{OV} = 0.23\ \text{V}$$

Thus,

$$v_{GS} = V_t + v_{OV} = 0.73\ \text{V}$$

and

$$v_{DS\text{min}} = v_{OV} = 0.23\ \text{V}$$

Example 5.1 *continued*

(c) For the MOSFET in the triode region with v_{DS} very small,

$$r_{DS} = \frac{1}{k_n' \frac{W}{L} v_{OV}}$$

Thus

$$1000 = \frac{1}{388 \times 10^{-6} \times 10 \times v_{OV}}$$

which gives

$$v_{OV} = 0.26 \text{ V}$$

Thus,

$$v_{GS} = 0.76 \text{ V}$$

EXERCISES

5.2 For a 65-nm process technology for which $t_{ox} = 1.4$ nm, $\mu_n = 216$ cm^2/V · s, and $V_t = 0.35$ V, find C_{ox}, k_n', and the values of v_{OV} and v_{GS} required to operate a transistor having $W/L = 10$ in saturation with $i_D = 50$ μA. What is the minimum value of v_{DS} needed?
Ans. 24.6 fF/μm^2; 531 μA/V^2; 0.14 V; 0.49 V; 0.14 V

D5.3 A circuit designer intending to operate a MOSFET in saturation is considering the effect of changing the device dimensions and operating voltages on the drain current i_D. Specifically, by what factor does i_D change in each of the following cases?
(a) The channel length is doubled.
(b) The channel width is doubled.
(c) The overdrive voltage is doubled.
(d) The drain-to-source voltage is doubled.
(e) Changes (a), (b), (c), and (d) are made simultaneously.
Which of these cases might cause the MOSFET to leave the saturation region?
Ans. 0.5; 2; 4; no change; 4; case (c) if v_{DS} is smaller than $2v_{OV}$

5.1.7 The *p*-Channel MOSFET

Figure 5.9(a) shows a cross section of a *p*-channel enhancement-type MOSFET. The structure is similar to that of the NMOS device except that here the substrate is *n* type and the source and the drain regions are p^+ type; that is, all semiconductor regions are reversed in polarity

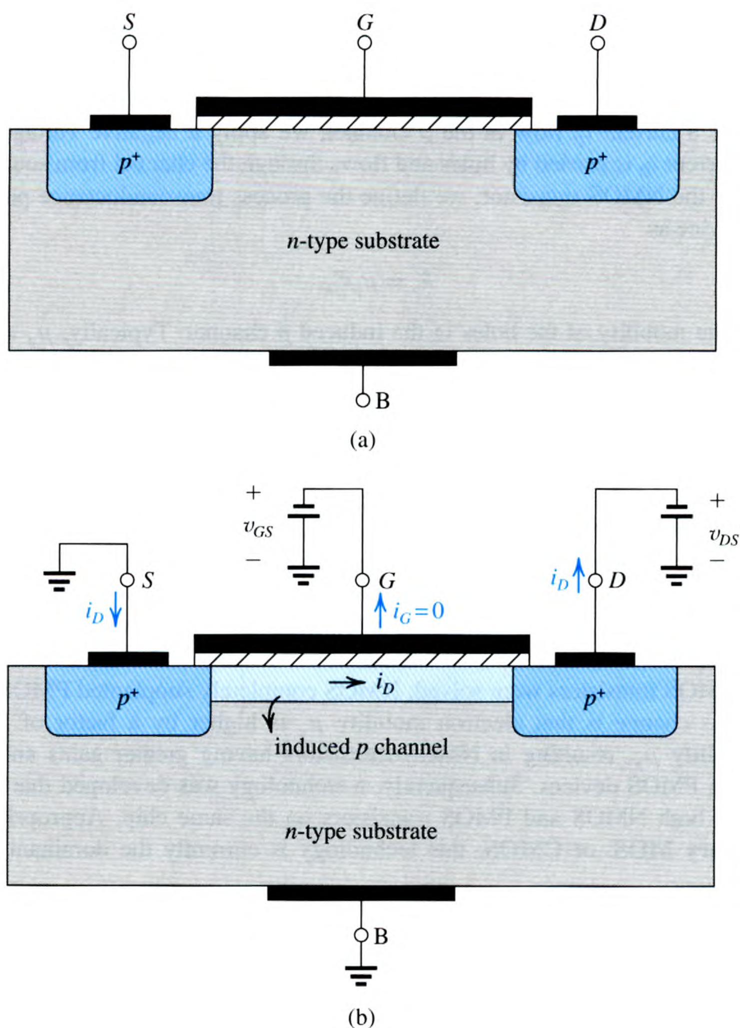


Figure 5.9 **(a)** Physical structure of the PMOS transistor. Note that it is similar to the NMOS transistor shown in Fig. 5.1(b) except that all semiconductor regions are reversed in polarity. **(b)** A negative voltage v_{GS} of magnitude greater than $|V_{tp}|$ induces a p channel, and a negative v_{DS} causes a current i_D to flow from source to drain.

relative to their counterparts in the NMOS case. The PMOS and NMOS transistors are said to be *complementary* devices.

To induce a channel for current flow between source and drain, a negative voltage is applied to the gate, that is, between gate and source, as indicated in Fig. 5.9(b). By increasing the magnitude of the negative v_{GS} beyond the magnitude of the threshold voltage V_{tp}, which by convention is negative, a p channel is established as shown in Fig. 5.9(b). This condition can be described as

$$v_{GS} \leq V_{tp}$$

or, to avoid dealing with negative signs,

$$|v_{SG}| \geq |V_{tp}|$$

Now, to make a current i_D flow in the p channel, we apply a negative voltage v_{DS} to the drain.[6] The current i_D is carried by holes and flows through the channel from source to drain. As we did for the NMOS transistor, we define the process transconductance parameter for the PMOS device as

$$k_p' = \mu_p C_{ox}$$

where μ_p is the mobility of the holes in the induced p channel. Typically, $\mu_p = 0.25\,\mu_n$ to $0.5\,\mu_n$ and is process-technology dependent. The transistor transconductance parameter k_p is obtained by multiplying k_p' by the aspect ratio W/L,

$$k_p = k_p'(W/L)$$

The remainder of the description of the physical operation of the p-channel MOSFET follows that for the NMOS device, except of course for the sign reversals of all voltages. We will present the complete current–voltage characteristics of both NMOS and PMOS transistors in the next section.

PMOS technology originally dominated MOS integrated-circuit manufacturing, and the original microprocessors utilized PMOS transistors. As the technological difficulties of fabricating NMOS transistors were solved, NMOS completely supplanted PMOS. The main reason for this change is that electron mobility μ_n is higher by a factor of 2 to 4 than the hole mobility μ_p, resulting in NMOS transistors having greater gains and speeds of operation than PMOS devices. Subsequently, a technology was developed that permits the fabrication of both NMOS and PMOS transistors on the same chip. Appropriately called **complementary MOS**, or **CMOS**, this technology is currently the dominant electronics technology.

5.1.8 Complementary MOS or CMOS

As the name implies, complementary MOS technology uses MOS transistors of both polarities. Although CMOS circuits are somewhat more difficult to fabricate than NMOS, the availability of complementary devices makes many powerful circuit configurations possible. Indeed, CMOS is now the most widely used of all the IC technologies in both analog and digital circuits. CMOS technology has virtually replaced designs based on NMOS transistors alone, and by 2019 CMOS technology had taken over many applications that just a few years earlier were possible only with bipolar devices. Throughout this book, we will study many CMOS circuit techniques.

Figure 5.10 shows a cross section of a CMOS chip illustrating how the PMOS and NMOS transistors are fabricated. Observe that while the NMOS transistor is built directly in the p-type substrate, the PMOS transistor is built in a specially created n region, known as an ***n* well**. The two devices are isolated from each other by a thick region of oxide that functions as an insulator.

---

[6]If a positive voltage is applied to the drain, the pn junction between the drain region and the substrate will become forward biased, and the device will no longer operate as a MOSFET. Proper MOSFET operation is predicated on the pn junctions between the source and drain regions and the substrate being always reverse biased.

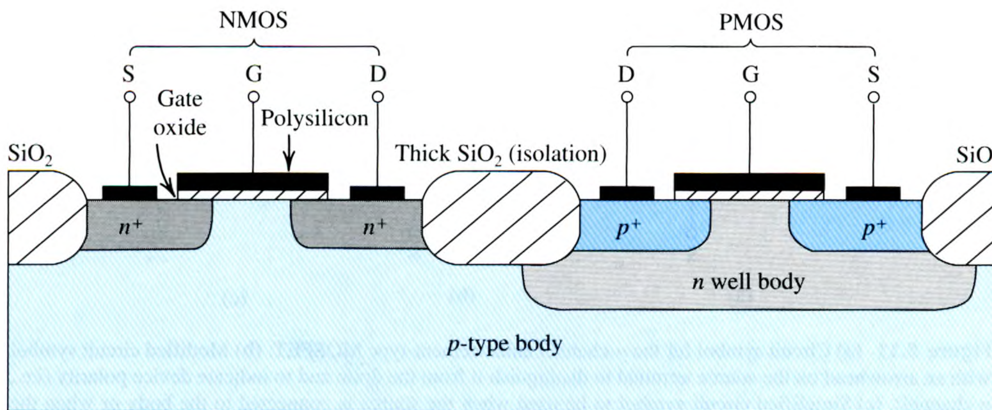


Figure 5.10 Cross section of a CMOS integrated circuit. Note that the PMOS transistor is formed in a separate n-type region, known as an n well, which functions as the body for the PMOS transistor. Not shown are the connections made to the p-type body and to the n well body.

5.2 Current–Voltage Characteristics

Building on the physical foundation established in the previous section for the operation of the enhancement MOS transistor, in this section we present its complete current–voltage characteristics. These characteristics can be measured at dc or at low frequencies and thus are called static characteristics. The dynamic effects that limit the operation of the MOSFET at high frequencies and high switching speeds will be discussed in Chapter 10.

5.2.1 Circuit Symbol

Figure 5.11(a) shows the circuit symbol for the n-channel enhancement-type MOSFET. The spacing between the two vertical lines that represent the gate and the channel indicates that the gate electrode is insulated from the body of the device. The polarity of the p-type substrate (body) and the n channel is indicated by the arrowhead on the line representing the body (B). This arrowhead also indicates the polarity of the transistor, namely, that it is an n-channel device.

Although the MOSFET is a symmetrical device, it is often useful in circuit design to designate one terminal as the source and the other as the drain (without having to write S and D beside the terminals). The modified circuit symbol in Fig. 5.11(b) accomplishes that by placing an arrowhead on the source terminal to distinguish it from the drain terminal. The arrowhead points in the normal direction of current flow and thus indicates the polarity of the device (i.e., n channel). Notice that in the modified symbol, there is no need to show the arrowhead on the body line. Although the circuit symbol of Fig. 5.11(b) clearly distinguishes the source from the drain, in practice it is the polarity of the voltage applied across the device that determines source and drain; *the drain is always positive relative to the source in an n-channel MOSFET.*

In applications where the source is connected to the body of the device, a further simplification of the circuit symbol is possible, as indicated in Fig. 5.11(c). This symbol is also used in applications when the effect of the body on circuit operation is not important, as will be seen later.

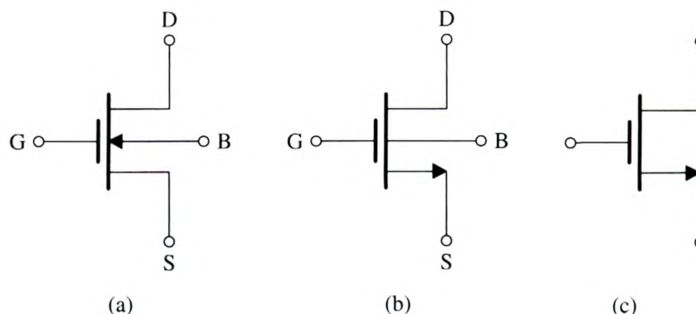


Figure 5.11 **(a)** Circuit symbol for the n-channel enhancement-type MOSFET. **(b)** Modified circuit symbol with an arrowhead on the source terminal to distinguish it from the drain and to indicate device polarity (i.e., n channel). **(c)** Simplified circuit symbol to be used when the source is connected to the body or when the effect of the body on device operation is unimportant.

5.2.2 The i_D–v_{DS} Characteristics

Table 5.1 provides the conditions and the formulas for the operation of the NMOS transistor in each of the three possible regions: the cutoff region, the triode region, and the saturation region. The cutoff and triode regions are useful when we want to operate the MOSFET as a switch. If we want to use the MOSFET to design an amplifier, it must be operated in the saturation region. The rationale for these choices will be addressed in Chapter 7.

At the top of Table 5.1 we show a circuit consisting of an NMOS transistor and two dc supplies providing v_{GS} and v_{DS}. This conceptual circuit can be used to measure the i_D–v_{DS} characteristic curves of the NMOS transistor. Each curve is measured by setting v_{GS} to a desired constant value, varying v_{DS}, and measuring the corresponding i_D. Two of these characteristic curves are shown in the accompanying diagram: one for $v_{GS} < V_{tn}$ and the other for $v_{GS} = V_{tn} + v_{OV}$. (Note that we are now using V_{tn} to denote the threshold voltage of the NMOS transistor, to distinguish it from that of the PMOS transistor, denoted V_{tp}.)

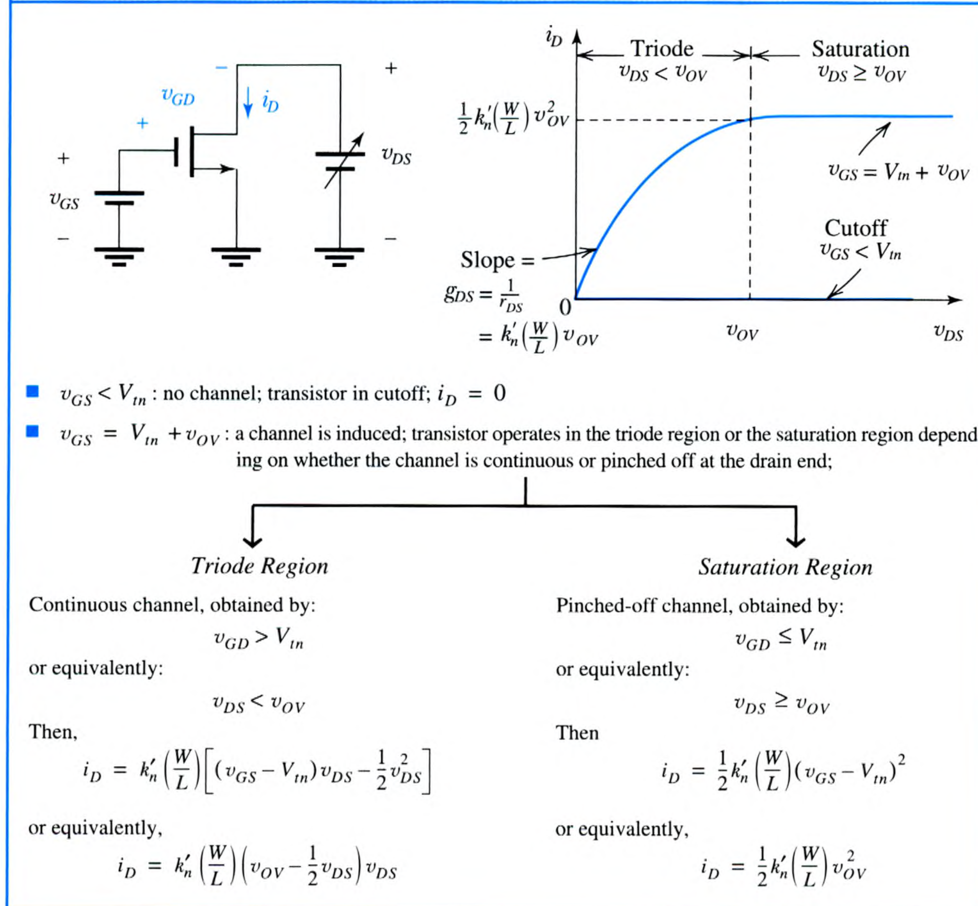
As Table 5.1 shows, the boundary between the triode region and the saturation region is determined by whether v_{DS} is less or greater than the overdrive voltage v_{OV} at which the transistor is operating. An equivalent way to check for the region of operation is to examine the relative values of the drain and gate voltages. To operate in the triode region, the gate voltage must exceed the drain voltage by at least V_{tn} volts, which ensures that the channel remains continuous (not pinched off). On the other hand, to operate in saturation, the channel must be pinched off at the drain end; pinch-off is achieved here by keeping v_D higher than $v_G - V_{tn}$, that is, not allowing v_D to fall below v_G by more than V_{tn} volts. Refer to Fig. 5.12 for a graphical reminder of these conditions.

Figure 5.13 shows a set of $i_D - v_{DS}$ characteristics for the NMOS transistor. Notice that each graph is obtained by setting v_{GS} above V_{tn} by a specific value of overdrive voltage, denoted v_{OV1}, v_{OV2}, v_{OV3}, and v_{OV4}. This in turn is the value of v_{DS} at which the corresponding graph *saturates*, and the value of the resulting saturation current is directly determined by the value of v_{OV}, namely, $\frac{1}{2}k_n v_{OV1}^2$, $\frac{1}{2}k_n v_{OV2}^2, \ldots$ You should memorize both the structure of these graphs and the coordinates of the saturation points.

Finally, notice that the boundary between the triode and the saturation regions, that is, the locus of the saturation points, is a parabolic curve described by

$$i_D = \frac{1}{2}k_n'\left(\frac{W}{L}\right)v_{DS}^2 \tag{5.20}$$

Table 5.1 Regions of Operation of the Enhancement NMOS Transistor



- $v_{GS} < V_{tn}$: no channel; transistor in cutoff; $i_D = 0$
- $v_{GS} = V_{tn} + v_{OV}$: a channel is induced; transistor operates in the triode region or the saturation region depending on whether the channel is continuous or pinched off at the drain end;

Triode Region

Continuous channel, obtained by:

$$v_{GD} > V_{tn}$$

or equivalently:

$$v_{DS} < v_{OV}$$

Then,

$$i_D = k_n'\left(\frac{W}{L}\right)\left[(v_{GS} - V_{tn})v_{DS} - \frac{1}{2}v_{DS}^2\right]$$

or equivalently,

$$i_D = k_n'\left(\frac{W}{L}\right)\left(v_{OV} - \frac{1}{2}v_{DS}\right)v_{DS}$$

Saturation Region

Pinched-off channel, obtained by:

$$v_{GD} \leq V_{tn}$$

or equivalently:

$$v_{DS} \geq v_{OV}$$

Then

$$i_D = \frac{1}{2}k_n'\left(\frac{W}{L}\right)(v_{GS} - V_{tn})^2$$

or equivalently,

$$i_D = \frac{1}{2}k_n'\left(\frac{W}{L}\right)v_{OV}^2$$

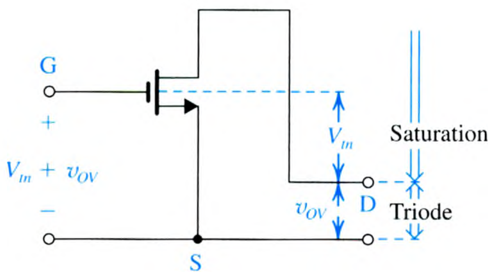


Figure 5.12 The relative levels of the terminal voltages of the enhancement NMOS transistor for operation in the triode region and in the saturation region.

5.2.3 The i_D–v_{GS} Characteristic

When the MOSFET is used to design an amplifier, it is operated in the saturation region. As Fig. 5.13 indicates, in saturation the drain current is a constant determined by v_{GS} (or v_{OV}) and is independent of v_{DS}. That is, the MOSFET operates as a constant-current source where

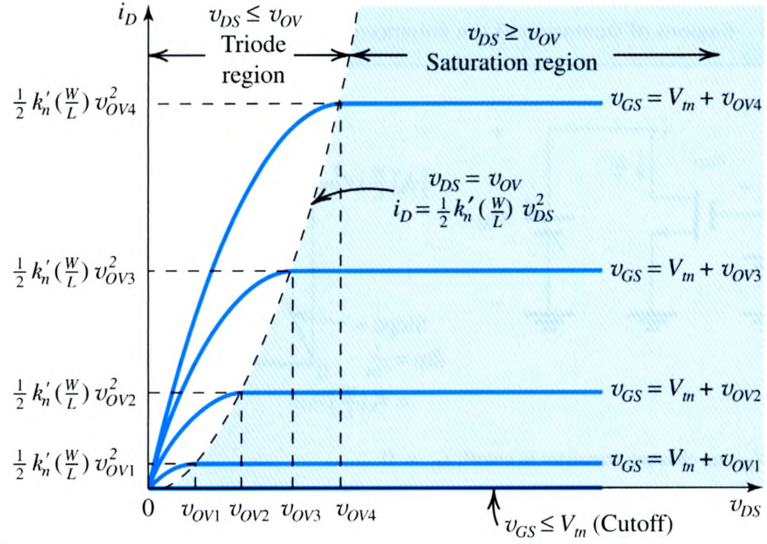


Figure 5.13 The i_D–v_{DS} characteristics for an enhancement-type NMOS transistor.

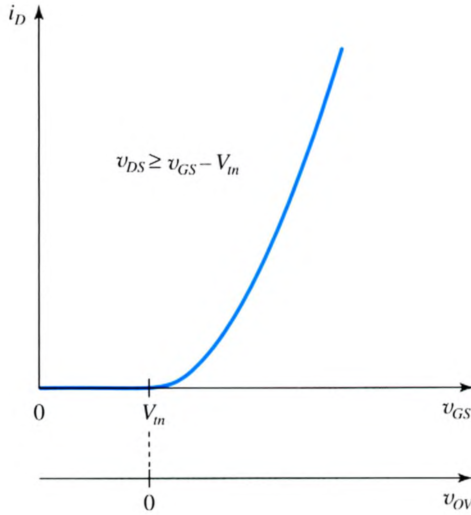


Figure 5.14 The i_D–v_{GS} characteristic of an NMOS transistor operating in the saturation region. The i_D–v_{OV} characteristic can be obtained by simply relabeling the horizontal axis, that is, shifting the origin to the point $v_{GS} = V_{tn}$.

the value of the current is determined by v_{GS}. In effect, then, the MOSFET operates as a voltage-controlled current source with the control relationship described by

$$i_D = \frac{1}{2} k_n' \left(\frac{W}{L}\right) (v_{GS} - V_{tn})^2 \tag{5.21}$$

or in terms of v_{OV},

$$i_D = \frac{1}{2} k_n' \left(\frac{W}{L}\right) v_{OV}^2 \tag{5.22}$$

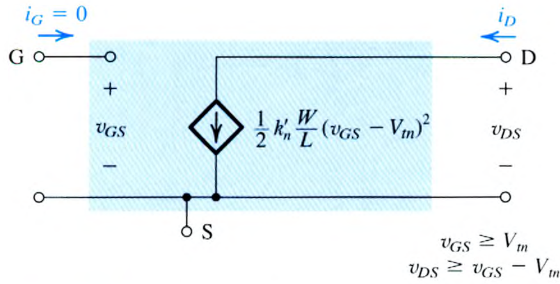


Figure 5.15 Large-signal, equivalent-circuit model of an n-channel MOSFET operating in the saturation region.

This is the relationship that underlies the application of the MOSFET as an amplifier. The fact that it is nonlinear creates a challenge when designing linear amplifiers, but we will show how to do this in Chapter 7.

Figure 5.14 shows the i_D–v_{GS} characteristic of an NMOS transistor operating in saturation. Note that if we want to plot i_D versus v_{OV}, we simply shift the origin to the point $v_{GS} = V_{tn}$.

The view of the MOSFET in the saturation region as a voltage-controlled current source is illustrated by the equivalent-circuit representation shown in Fig. 5.15. For reasons that will become apparent shortly, the circuit in Fig. 5.15 is known as a **large-signal equivalent circuit**. Note that the current source is ideal,[7] with an infinite output resistance representing the independence, in saturation, of i_D from v_{DS}. This is what we have assumed in the idealized model of device operation we have used so far. We are about to rectify an important shortcoming of this model. First, however, we present an example.

Example 5.2

Consider an NMOS transistor fabricated in a 0.18-μm process with $L = 0.18$ μm and $W = 2$ μm. The process technology is specified to have $C_{ox} = 8.6$ fF/μm^2, $\mu_n = 450$ cm^2/V·s, and $V_{tn} = 0.5$ V.

(a) Find v_{GS} and v_{DS} that result in the MOSFET operating at the edge of saturation with $i_D = 100$ μA.
(b) If v_{GS} is kept constant, find v_{DS} that results in $i_D = 50$ μA.

Solution

First we determine the process transconductance parameter k_n',

$$k_n' = \mu_n C_{ox}$$
$$= 450 \times 10^{-4} \times 8.6 \times 10^{-15} \times 10^{12} \text{ A/V}^2 = 387 \text{ μA/V}^2$$

and the transistor transconductance parameter k_n,

$$k_n = k_n'\left(\frac{W}{L}\right) = 387\left(\frac{2}{0.18}\right) = 4.3 \text{ mA/V}^2$$

[7]An ideal current source is one whose current is independent of the voltage across its terminals. Thus, its output resistance must be infinite.

Example 5.2 *continued*

(a) With the transistor operating in saturation,

$$i_D = \frac{1}{2}k_n v_{OV}^2$$

Thus,

$$100 = \frac{1}{2} \times 4.3 \times 10^3 \times v_{OV}^2$$

which gives us

$$v_{OV} = 0.22 \text{ V}$$

Thus,

$$v_{GS} = V_{tn} + v_{OV} = 0.5 + 0.22 = 0.72 \text{ V}$$

and since operation is at the edge of saturation,

$$v_{DS} = v_{OV} = 0.22 \text{ V}$$

(b) With v_{GS} kept constant at 0.72 V, and i_D reduced from the value obtained at the edge of saturation, the MOSFET will now be operating in the triode region, thus:

$$i_D = k_n\left[v_{OV}v_{DS} - \frac{1}{2}v_{DS}^2\right]$$

$$50 = 4.3 \times 10^3\left[0.22v_{DS} - \frac{1}{2}v_{DS}^2\right]$$

which can be rearranged to the form

$$v_{DS}^2 - 0.44v_{DS} + 0.023 = 0$$

This quadratic equation has two solutions

$$v_{DS} = 0.06 \text{ V} \quad \text{and} \quad v_{DS} = 0.39 \text{ V}$$

The second answer is greater than v_{OV}. Since we know that the transistor is operating in the triode region, we can disregard it. Thus we have

$$v_{DS} = 0.06 \text{ V}$$

EXERCISES

5.4 An NMOS transistor is operating at the edge of saturation with an overdrive voltage v_{OV} and a drain current i_D. If v_{OV} is doubled, and we must maintain operation at the edge of saturation, what should v_{DS} be changed to? What value of drain current results?
Ans. $2v_{OV}$; $4i_D$

5.5 An n-channel MOSFET operating with $v_{OV} = 0.5$ V exhibits a linear resistance $r_{DS} = 1$ kΩ when v_{DS} is very small. What is the value of the device transconductance parameter k_n? What is the value of the current i_D obtained when v_{DS} is increased to 0.2 V? to 0.5 V? and to 1 V?
Ans. 2 mA/V^2; 0.16 mA; 0.25 mA; 0.25 mA

5.2.4 Finite Output Resistance in Saturation

Equation (5.21) and the corresponding large-signal equivalent circuit in Fig. 5.15, as well as the graphs in Fig. 5.13, show that in saturation, i_D is independent of v_{DS}. Thus, a change Δv_{DS} in the drain-to-source voltage causes a zero change in i_D, which implies that the incremental resistance looking into the drain of a saturated MOSFET is infinite. This, however, is an idealization based on the premise that once the channel is pinched off at the drain end, further increases in v_{DS} have no effect on the channel's shape. But, in practice, increasing v_{DS} beyond v_{OV} does affect the channel somewhat. As v_{DS} is increased, the channel pinch-off point moves slightly away from the drain, toward the source. This is illustrated in Fig. 5.16, which shows that the voltage across the channel remains constant at v_{OV}, and the additional voltage applied to the drain appears as a voltage drop across the narrow depletion region between the end of the channel and the drain region. This voltage accelerates the electrons that reach the drain end of the channel and sweeps them across the depletion region into the drain. Note, however, that (with depletion-layer widening) the channel length is in effect reduced, from L to $L - \Delta L$, a phenomenon known as **channel-length modulation**. Now, since i_D is inversely proportional to the channel length (Eq. 5.21), i_D increases with v_{DS}.

We account for this effect in the expression for i_D by including a factor $1 + \lambda(v_{DS} - v_{OV})$ or, for simplicity, $(1 + \lambda v_{DS})$,

$$i_D = \frac{1}{2}k_n'\left(\frac{W}{L}\right)(v_{GS} - V_{tn})^2(1 + \lambda v_{DS}) \tag{5.23}$$

Here λ is a device parameter whose units are reciprocal volts (V^{-1}). The value of λ depends both on the process technology used to fabricate the device and on the channel length L that the circuit designer selects. Specifically, the value of λ is much larger for newer submicron technologies than for older technologies. This makes intuitive sense: Newer technologies have very short channels, and are thus much more greatly impacted by the channel-length modulation effect. Also, for a given process technology, λ is inversely proportional to L.

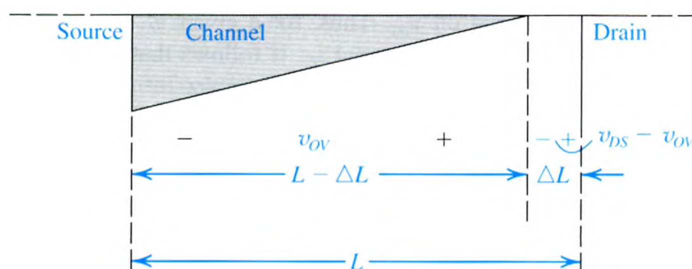


Figure 5.16 Increasing v_{DS} beyond v_{DSsat} (which is equal to v_{OV}) causes the channel pinch-off point to move slightly away from the drain, thus reducing the effective channel length (by ΔL).

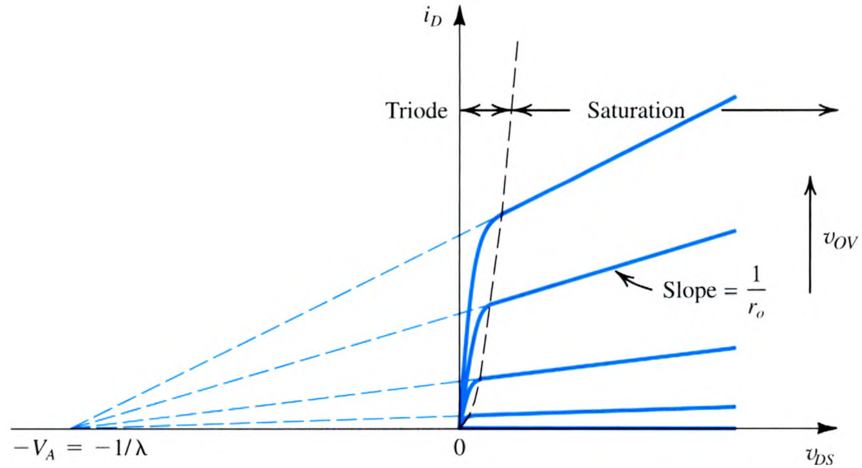


Figure 5.17 Effect of v_{DS} on i_D in the saturation region. The MOSFET parameter V_A depends on the process technology and, for a given process, is proportional to the channel length L.

Figure 5.17 presents a typical set of i_D–v_{DS} characteristics showing the effect of channel-length modulation. The observed linear dependence of i_D on v_{DS} in the saturation region is represented in Eq. (5.23) by the factor $(1 + \lambda v_{DS})$. From Fig. 5.17 we can see that when the straight-line i_D–v_{DS} characteristics are extrapolated, they intercept the v_{DS} axis at the point, $v_{DS} = -V_A$, where V_A is a positive voltage. Equation (5.23), however, indicates that $i_D = 0$ at $v_{DS} = -1/\lambda$. It follows that

$$V_A = \frac{1}{\lambda}$$

and thus V_A is a device parameter with the dimensions of V. For a given process, V_A is proportional to the channel length L that the designer selects for a MOSFET. We can isolate the dependence of V_A on L by expressing it as

$$V_A = V_A' L$$

where V_A' is entirely process-technology dependent, with the dimensions of volts per micron. Typically, V_A' falls in the range of 5 V/μm to 50 V/μm. The voltage V_A is usually referred to as the Early voltage, after J. M. Early, who discovered a similar phenomenon for the BJT (See Chapter 6).

Equation (5.23) indicates that when channel-length modulation is taken into account, the saturation values of i_D depend on v_{DS}. Thus, for a given v_{GS}, a change Δv_{DS} yields a corresponding change Δi_D in the drain current i_D. It follows that the output resistance of the current source representing i_D in saturation is no longer infinite. Defining the output resistance r_o as[8]

$$r_o \equiv \left[\frac{\partial i_D}{\partial v_{DS}}\right]^{-1}_{v_{GS}\text{ constant}} \tag{5.24}$$

[8]In this book we use r_o to denote the output resistance in saturation, and r_{DS} to denote the drain-to-source resistance in the triode region, for small v_{DS}.

and using Eq. (5.23) results in

$$r_o = \left[\lambda \frac{k_n'}{2} \frac{W}{L} (v_{GS} - V_{tn})^2 \right]^{-1} \tag{5.25}$$

which can be written as

$$r_o = \frac{1}{\lambda i_D'} \tag{5.26}$$

or, equivalently,

$$r_o = \frac{V_A}{i_D'} \tag{5.27}$$

where i_D' is the drain current *without* channel-length modulation taken into account; that is,

$$i_D' = \frac{1}{2} k_n' \frac{W}{L} (v_{GS} - V_{tn})^2 \tag{5.27'}$$

Thus the output resistance is inversely proportional to the drain current.[9] Finally, we show in Fig. 5.18 the large-signal, equivalent-circuit model incorporating r_o.

EXERCISE

5.6 An NMOS transistor is fabricated in a 0.18-μm process having $\mu_n C_{ox} = 400\ \mu\text{A/V}^2$ and $V_A' = 5\ \text{V/}\mu\text{m}$ of channel length. If $L = 0.8\ \mu\text{m}$ and $W = 16\ \mu\text{m}$, find V_A and λ. Find the value of i_D that results when the device is operated with an overdrive voltage $v_{OV} = 0.2$ V and $v_{DS} = 0.8$ V. Also, find the value of r_o at this operating point. If v_{DS} is increased by 1 V, what is the corresponding change in i_D?
Ans. 4 V; 0.25 V^{-1}; 192 μA; 25 kΩ; 40 μA

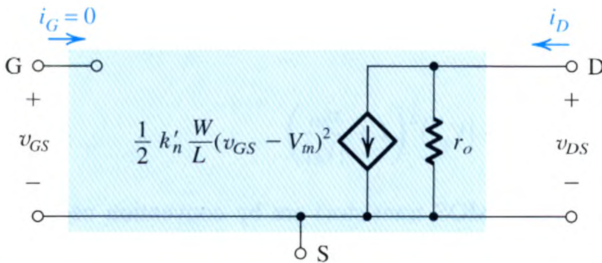


Figure 5.18 Large-signal, equivalent-circuit model of the *n*-channel MOSFET in saturation, incorporating the output resistance r_o. The output resistance models the linear dependence of i_D on v_{DS} and is given by Eq. (5.27).

[9] In applying Eq. (5.27) we will usually drop the prime on i_D and simply use $r_o = V_A/i_D$ where i_D is the drain current without channel-length modulation.

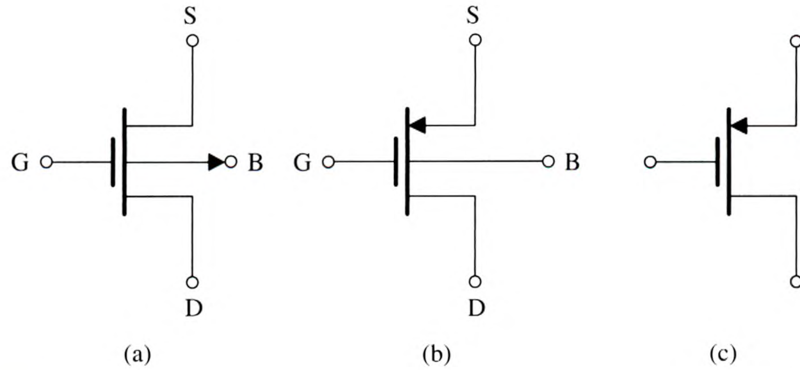


Figure 5.19 **(a)** Circuit symbol for the p-channel enhancement-type MOSFET. **(b)** Modified symbol with an arrowhead on the source lead. **(c)** Simplified symbol for cases where the source is connected to the body.

5.2.5 Characteristics of the *p*-Channel MOSFET

The circuit symbol for the p-channel enhancement-type MOSFET is shown in Fig. 5.19(a). Figure 5.19(b) shows a modified circuit symbol in which an arrowhead pointing in the normal direction of current flow is included on the source terminal. When the source is connected to the substrate, the simplified symbol of Fig. 5.19(c) is usually used.

The regions of operation of the PMOS transistor and the corresponding conditions and expression for i_D are shown in Table 5.2. Note that the equations are written in a way that emphasizes physical intuition and avoids the confusion of negative signs. Thus while V_{tp} is by convention negative, we use $|V_{tp}|$, and the voltages v_{SG} and v_{SD} are positive. Also, in all of our circuit diagrams we will always draw p-channel devices with their sources on top so that current flows from top to bottom. Finally, we note that PMOS devices also suffer from the channel-length modulation effect. This can be taken into account by including a factor $(1+|\lambda|v_{SD})$ in the saturation-region expression for i_D as follows

$$i_D = \frac{1}{2}k_p'\left(\frac{W}{L}\right)\left(v_{SG}-|V_{tp}|\right)^2(1+|\lambda|v_{SD}) \tag{5.28}$$

or equivalently

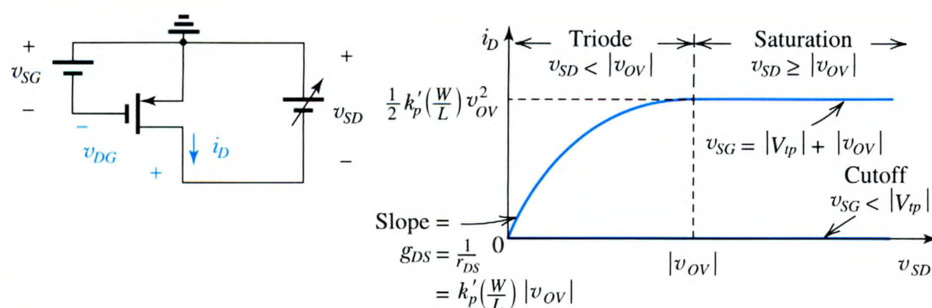
$$i_D = \frac{1}{2}k_p'\left(\frac{W}{L}\right)\left(v_{SG}-|V_{tp}|\right)^2\left(1+\frac{v_{SD}}{|V_A|}\right) \tag{5.29}$$

where λ and V_A (the Early voltage for the PMOS transistor) are by convention negative quantities, hence we use $|\lambda|$ and $|V_A|$.

Finally, we should note that for a given CMOS fabrication process λ_n and $|\lambda_p|$ are generally not equal, and similarly for V_{An} and $|V_{Ap}|$.

To recap, to turn a PMOS transistor on, the gate voltage has to be made lower than that of the source by at least $|V_{tp}|$. To operate in the triode region, the drain voltage has to exceed that of the gate by at least $|V_{tp}|$; otherwise, the PMOS operates in saturation. Finally, Fig. 5.20 provides a pictorial representation of these operating conditions.

Table 5.2 Regions of Operation of the Enhancement PMOS Transistor



- $v_{SG} < |V_{tp}|$: no channel; transistor in cutoff; $i_D = 0$
- $v_{SG} = |V_{tp}| + |v_{OV}|$: a channel is induced; transistor operates in the triode region or in the saturation region depending on whether the channel is continuous or pinched off at the drain end;

Triode Region	*Saturation Region*
Continuous channel, obtained by:	Pinched-off channel, obtained by:
$v_{DG} > \lvert V_{tp}\rvert$	$v_{DG} \leq \lvert V_{tp}\rvert$
or equivalently	or equivalently
$v_{SD} < \lvert v_{OV}\rvert$	$v_{SD} \geq \lvert v_{OV}\rvert$
Then	Then
$i_D = k_p'\left(\frac{W}{L}\right)\left[(v_{SG} - \lvert V_{tp}\rvert)v_{SD} - \frac{1}{2}v_{SD}^2\right]$	$i_D = \frac{1}{2}k_p'\left(\frac{W}{L}\right)(v_{SG} - \lvert V_{tp}\rvert)^2$
or equivalently	or equivalently
$i_D = k_p'\left(\frac{W}{L}\right)\left(\lvert v_{OV}\rvert - \frac{1}{2}v_{SD}\right)v_{SD}$	$i_D = \frac{1}{2}k_p'\left(\frac{W}{L}\right)v_{OV}^2$

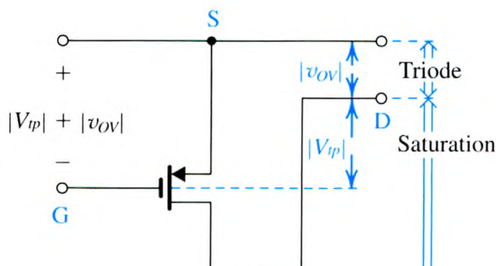


Figure 5.20 The relative levels of the terminal voltages of the enhancement-type PMOS transistor for operation in the triode region and in the saturation region.

Video Example VE 5.1

The PMOS transistor in Fig. VE5.1 has $V_{tp} = -0.5$ V. As the gate voltage v_G is varied from +3 V to 0 V, the transistor moves through all of its three possible modes of operation. Specify the values of v_G at which the device changes modes of operation.

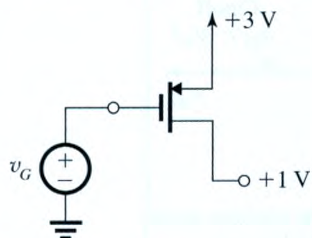


Figure VE5.1 Circuit for Video Example 5.1.

VE 5.1

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 5.40

EXERCISE

5.7 The PMOS transistor shown in Fig. E5.7 has $V_{tp} = -0.5$ V, $k_p' = 100\ \mu\text{A/V}^2$, and $W/L = 10$.

(a) Find the range of v_G for which the transistor conducts.

(b) In terms of v_G, find the range of v_D for which the transistor operates in the triode region.

(c) In terms of v_G, find the range of v_D for which the transistor operates in saturation.

(d) Neglecting channel-length modulation (i.e., assuming $\lambda = 0$), find the values of $|v_{OV}|$ and v_G and the corresponding range of v_D to operate the transistor in the saturation mode with $i_D = 50\ \mu$A.

(e) If $\lambda = -0.2\ \text{V}^{-1}$, find the value of r_o corresponding to the overdrive voltage determined in (d).

(f) For $\lambda = -0.2\ \text{V}^{-1}$ and for the value of v_{OV} determined in (d), find i_D at $v_D = +1$ V and at $v_D = 0$ V; hence, calculate the value of the apparent output resistance in saturation. Compare to the value found in (e).

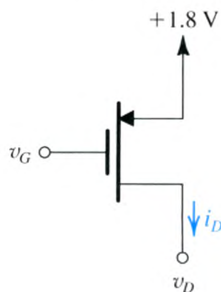


Figure E5.7

Ans. (a) $v_G \leq +1.3$ V; (b) $v_D \geq v_G + 0.5$; (c) $v_D \leq V_G + 0.5$; (d) 0.32 V, 0.98 V, $v_D \leq 1.48$ V; (e) 100 kΩ; (f) 58 μA, 68 μA, 100 kΩ (same)

5.3 MOSFET Circuits at DC

Having studied the current–voltage characteristics of MOSFETs, we now consider circuits involving only dc voltages and currents. Specifically, we shall present a series of design and analysis examples of MOSFET circuits operating at dc. The goal of these examples is to make you familiar with the device and to help you acquire the ability to perform MOSFET circuit analysis both rapidly and effectively.

To keep matters simple and focus our attention on the essence of MOSFET circuit operation, we will generally neglect channel-length modulation; that is, we will assume $\lambda = 0$. We will find it convenient to work in terms of the overdrive voltage; $V_{OV} = V_{GS} - V_{tn}$ for NMOS and $|V_{OV}| = V_{SG} - |V_{tp}|$ for PMOS. Also, since we are only dealing with dc quantities, we will use uppercase symbols with uppercase subscripts for all currents and voltages.

Example 5.3

Design the circuit of Fig. 5.21: that is, determine the values of R_D and R_S so that the transistor operates at $I_D = 0.2$ mA and $V_D = +0.2$ V. The NMOS transistor has $V_t = 0.5$ V, $\mu_n C_{ox} = 400\ \mu\text{A/V}^2$, $L = 0.5\ \mu$m, and $W = 15\ \mu$m. Neglect the channel-length modulation effect (i.e., assume that $\lambda = 0$).

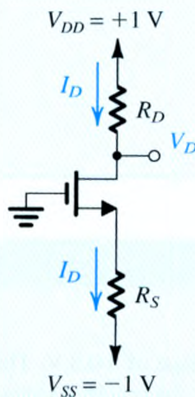


Figure 5.21 Circuit for Example 5.3.

Example 5.3 *continued*

Solution

To establish a dc voltage of +0.2 V at the drain, we must select R_D as follows:

$$R_D = \frac{V_{DD} - V_D}{I_D} = \frac{1 - 0.2}{0.2} = 4\ \text{k}\Omega$$

To determine the value required for R_S, we need to know the voltage at the source, which can be easily found if we know V_{GS}. This in turn can be determined from V_{OV}. Toward that end, we note that since $V_D = 0.2$ V is greater than V_G, the NMOS transistor is operating in the saturation region, and we can use the saturation-region expression of i_D to determine the required value of V_{OV},

$$I_D = \frac{1}{2}\mu_n C_{ox} \frac{W}{L} V_{OV}^2$$

Then substituting $I_D = 0.2$ mA $= 200\ \mu$A, $\mu_n C_{ox} = 400\ \mu\text{A/V}^2$, and $W/L = 15/0.5$ gives

$$200 = \frac{1}{2} \times 400 \times \frac{15}{0.5} V_{OV}^2$$

which results in

$$V_{OV} = 0.18\ \text{V}$$

Thus,

$$V_{GS} = V_t + V_{OV} = 0.5 + 0.18 = 0.68\ \text{V}$$

Referring to Fig. 5.21, we note that the gate is at ground potential. Thus, the source must be at –0.68 V, and the required value of R_S can be determined from

$$R_S = \frac{V_S - V_{SS}}{I_D} = \frac{-0.68 - (-1)}{0.2} = 1.6\ \text{k}\Omega$$

Video Example VE 5.2

Design the circuit of Fig. 5.21 to establish a drain current of 0.1 mA and a drain voltage of +0.3 V. The MOSFET has $V_t = 0.5$ V, $\mu_n C_{ox} = 400\ \mu\text{A/V}^2$, $L = 0.4\ \mu$m, and $W = 5\ \mu$m. Specify the required values for R_S and R_D. Assume $\lambda = 0$.

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.


VE 5.2

Related end-of-chapter problem: 5.43

EXERCISE

D5.8 Redesign the circuit of Fig. 5.21 for the following case: $V_{DD} = -V_{SS} = 1$ V, $V_t = 0.4$ V, $\mu_n C_{ox} = 400\ \mu\text{A/V}^2$, $W/L = 5\ \mu\text{m}/0.4\ \mu\text{m}$, $I_D = 100\ \mu\text{A}$, and $V_D = +0.2$ V.
Ans. $R_D = 8\ \text{k}\Omega$; $R_S = 4\ \text{k}\Omega$

Example 5.4

Figure 5.22 shows an NMOS transistor with its drain and gate terminals connected together. Find the i–v relationship of the resulting two-terminal device in terms of the MOSFET parameters $k_n = k_n'(W/L)$ and V_{tn}. Neglect channel-length modulation (i.e., $\lambda = 0$). Note that this two-terminal device is known as a **diode-connected transistor**.

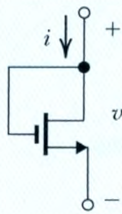


Figure 5.22 Circuit for Example 5.4.

Solution

We first determine which region the MOSFET is operating in. Since $v_D = v_G$, operation is in the saturation mode, thus

$$i_D = \frac{1}{2}k_n'\left(\frac{W}{L}\right)(v_{GS} - V_{tn})^2$$

Example 5.4 *continued*

Now, $i = i_D$ and $v = v_{GS}$, thus

$$i = \frac{1}{2}k_n'\left(\frac{W}{L}\right)(v - V_{tn})^2$$

Replacing $k_n'\left(\frac{W}{L}\right)$ by k_n results in

$$i = \frac{1}{2}k_n(v - V_{tn})^2$$

Video Example VE 5.3

The NMOS transistors in the circuit of Fig. VE5.3 have $V_t = 0.5$ V, $\mu_n C_{ox} = 250\ \mu\text{A/V}^2$, $\lambda = 0$, and $L_1 = L_2 = 0.25\ \mu\text{m}$. Find the required values of gate width for each of Q_1 and Q_2, and the value of R, to obtain the voltage and current values indicated.

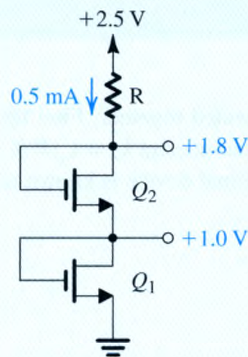


Figure VE5.3 Circuit for Video Example 5.3.


VE 5.3

Solution: Go to **www.oup.com/he/sedra-smith8e** to watch the authors solve this problem.

Related end-of-chapter problem: 5.49

EXERCISES

D5.9 For the circuit in Fig. E5.9, find the value of R that results in $V_D = 0.7$ V. The MOSFET has $V_{tn} = 0.5$ V, $\mu_n C_{ox} = 0.4\ \text{mA/V}^2$, $W/L = \frac{0.72\ \mu\text{m}}{0.18\ \mu\text{m}}$, and $\lambda = 0$.

Ans. 34.4 kΩ

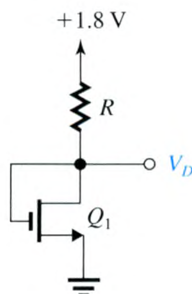


Figure E5.9

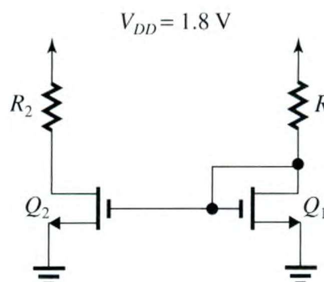


Figure E5.10

D5.10 Figure E5.10 shows a circuit obtained by augmenting the circuit of Fig. E5.9 with a transistor Q_2 identical to Q_1 and a resistance R_2. Find the value of R_2 that results in Q_2 operating at the edge of the saturation region. Use your solution to Exercise 5.9.
Ans. 50 kΩ

Example 5.5

Design the circuit in Fig. 5.23 to establish a drain voltage of 0.1 V. What is the effective resistance between drain and source at this operating point? Let $V_{tn} = 0.5$ V and $k_n'(W/L) = 2$ mA/ V^2.

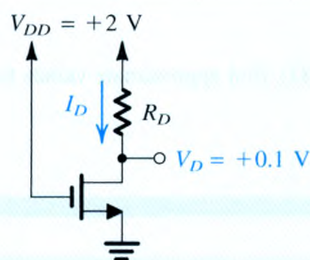


Figure 5.23 Circuit for Example 5.5.

Solution

Since the drain voltage is lower than the gate voltage by 1.9 V and $V_{tn} = 0.5$ V, the MOSFET is operating in the triode region. Thus the current I_D is given by

$$I_D = k_n' \frac{W}{L}\left[(V_{GS} - V_{tn})V_{DS} - \frac{1}{2}V_{DS}^2\right]$$

$$I_D = 2 \times \left[(2 - 0.5) \times 0.1 - \frac{1}{2} \times 0.01\right] = 0.29 \text{ mA}$$

Example 5.5 *continued*

The required value for R_D can be found as follows:

$$R_D = \frac{V_{DD} - V_D}{I_D} = \frac{2 - 0.1}{0.29} = 6.55 \text{ k}\Omega$$

In a practical discrete-circuit design problem, one selects the closest standard value available for, say, 5% resistors—in this case, 6.8 kΩ (see Appendix J). Since the transistor is operating in the triode region with a small V_{DS}, the effective drain-to-source resistance can be determined as follows:

$$r_{DS} = \frac{V_{DS}}{I_D} = \frac{0.1V}{0.29\text{mA}} = 345 \ \Omega$$

Alternatively, we can determine r_{DS} by using the formula

$$r_{DS} = \frac{1}{k_n V_{OV}}$$

to obtain

$$r_{DS} = \frac{1}{2 \times (2 - 0.5)} = 0.333 \text{ k}\Omega = 333 \ \Omega$$

which is close to the value found above.

EXERCISE

5.11 If in the circuit of Example 5.5 the value of R_D is doubled (to 13.1 kΩ), find approximate values for I_D and V_D.
Ans. 0.15 mA; 0.05 V

Example 5.6

Analyze the circuit shown in Fig. 5.24(a) to determine the voltages at all nodes and the currents through all branches. Let $V_{tn} = 1$ V and $k_n'(W/L) = 1$ mA/V^2. Neglect the channel-length modulation effect (i.e., assume $\lambda = 0$).

Solution

Since the gate current is zero, the voltage at the gate is simply determined by the voltage divider formed by the two 10-MΩ resistors,

$$V_G = V_{DD}\frac{R_{G2}}{R_{G2} + R_{G1}} = 10 \times \frac{10}{10 + 10} = +5 \text{ V}$$

Figure 5.24 **(a)** Circuit for Example 5.6. **(b)** The circuit with some of the analysis details shown.

With this positive voltage at the gate, the NMOS transistor will be turned on. We do not know, however, whether the transistor will be operating in the saturation region or in the triode region. We shall assume saturation-region operation, solve the problem, and then check the validity of our assumption. Obviously, if our assumption turns out not to be valid, we will have to solve the problem again for triode-region operation.

Refer to Fig. 5.24(b). Since the voltage at the gate is 5 V and the voltage at the source is I_D (mA) × 6 (kΩ) = $6I_D$ (V), we have

$$V_{GS} = 5 - 6I_D$$

Thus I_D is given by

$$I_D = \frac{1}{2}k_n'\frac{W}{L}(V_{GS} - V_{tn})^2$$
$$= \frac{1}{2} \times 1 \times (5 - 6I_D - 1)^2$$

which results in the following quadratic equation in I_D:

$$18I_D^2 - 25I_D + 8 = 0$$

This equation yields two values for I_D: 0.89 mA and 0.5 mA. The first value results in a source voltage of $6 \times 0.89 = 5.34$ V, which is greater than the gate voltage and does not make physical sense as it would imply that the NMOS transistor is cut off. Thus,

$$I_D = 0.5 \text{ mA}$$
$$V_S = 0.5 \times 6 = +3 \text{ V}$$
$$V_{GS} = 5 - 3 = 2 \text{ V}$$
$$V_D = 10 - 6 \times 0.5 = +7 \text{ V}$$

Since $V_D > V_G - V_{tn}$, the transistor is operating in saturation, as initially assumed.

EXERCISES

5.12 For the circuit of Fig. 5.24, what is the largest value that R_D can have while the transistor remains in the saturation mode?

Ans. 12 kΩ

D5.13 Redesign the circuit of Fig. 5.24 for the following requirements: $V_{DD} = +5$ V, $I_D = 0.32$ mA, $V_S = 1.6$ V, $V_D = 3.4$ V, with a 1-μA current through the voltage divider R_{G1}, R_{G2}. Assume the same MOSFET as in Example 5.6.

Ans. $R_{G1} = 1.6$ MΩ; $R_{G2} = 3.4$ MΩ; $R_S = R_D = 5$ kΩ

Example 5.7

Design the circuit of Fig. 5.25 so that the transistor operates in saturation with $I_D = 0.5$ mA and $V_D = +3$ V. Let the PMOS transistor have $V_{tp} = -1$ V and $k_p'(W/L) = 1$ mA/V^2. Assume $\lambda = 0$. What is the largest value that R_D can have while maintaining saturation-region operation?

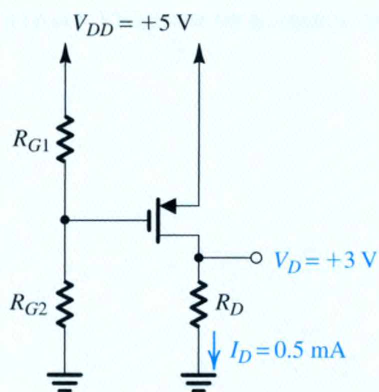


Figure 5.25 Circuit for Example 5.7.

Solution

Since the MOSFET is to be in saturation, we can write

$$I_D = \frac{1}{2}k_p'\frac{W}{L}|V_{OV}|^2$$

Substituting $I_D = 0.5$ mA and $k_p'W/L = 1$ mA/V^2, we obtain

$$|V_{OV}| = 1 \text{ V}$$

and

$$V_{SG} = |V_{tp}| + |V_{OV}| = 1 + 1 = 2 \text{ V}$$

Since the source is at +5 V, the gate voltage must be set to +3 V. We can do this by selecting appropriate values for R_{G1} and R_{G2}. To keep the current through R_{G1} and R_{G2} small, e.g., in the μA range, we select values for R_{G1} and R_{G2} in the MΩ range. A possible selection is $R_{G1} = 2$ MΩ and $R_{G2} = 3$ MΩ.

The value of R_D can be found from

$$R_D = \frac{V_D}{I_D} = \frac{3\text{ V}}{0.5\text{ mA}} = 6\text{ k}\Omega$$

To find the largest value that R_D can have while the MOSFET remains in saturation, recall that saturation-mode operation is maintained up to the point that V_D exceeds V_G by $|V_{tp}|$; that is, until

$$V_{D\max} = 3 + 1 = 4\text{ V}$$

This value of drain voltage is obtained with R_D given by

$$R_D = \frac{4\text{ V}}{0.5\text{ mA}} = 8\text{ k}\Omega$$

EXERCISE

D5.14 For the circuit in Fig. E5.14, find the value of R that casues the PMOS transistor to operate with an overdrive voltage $|V_{OV}| = 0.6$ V. The threshold voltage is $V_{tp} = -0.4$ V, the process transconductance parameter $k_p' = 0.1$ mA/V^2, and $W/L = 10$ μm/0.18 μm.

Ans. 800 Ω

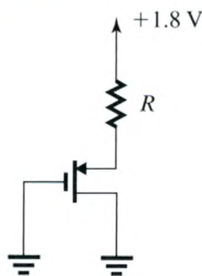


Figure E5.14

Example 5.8

The NMOS and PMOS transistors in the circuit of Fig. 5.26(a) are matched, with $k_n'(W_n/L_n) = k_p'(W_p/L_p) = 1$ mA/V^2 and $V_{tn} = -V_{tp} = 1$ V. Assuming $\lambda = 0$ for both devices, find the drain currents I_{DN} and I_{DP}, as well as the voltage V_O, for $V_I = 0$ V, +2.5 V, and −2.5 V.

Example 5.8 *continued*

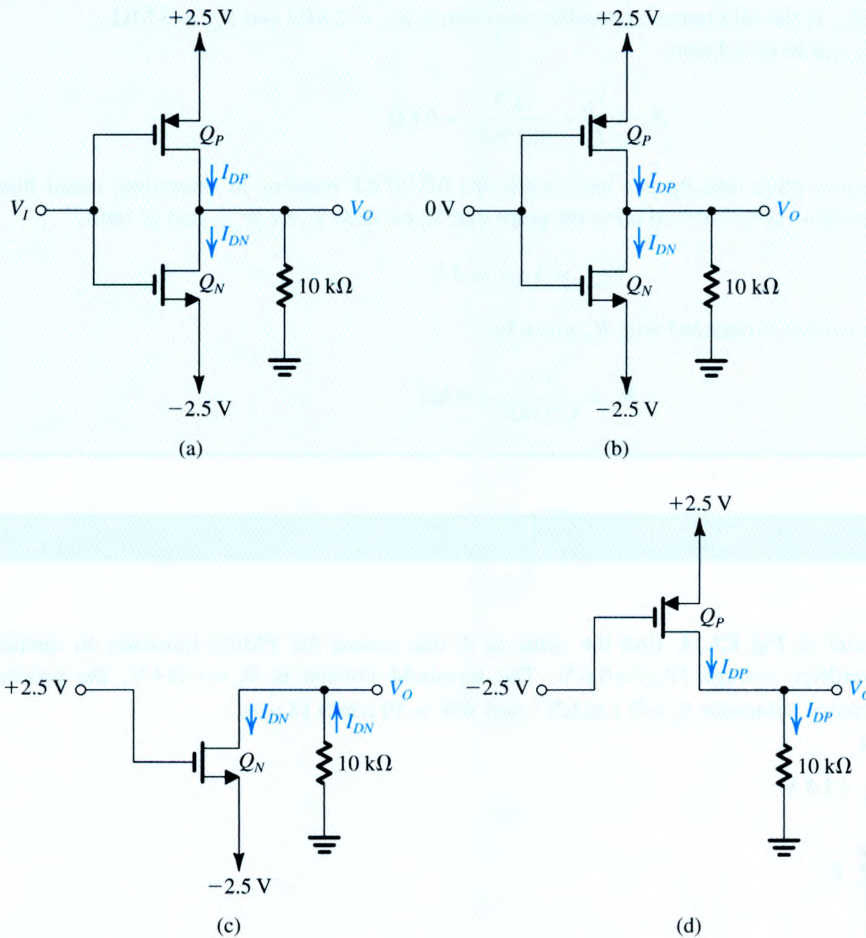


Figure 5.26 Circuits for Example 5.8.

Solution

Figure 5.26(b) shows the circuit for the case $V_I = 0$ V. We note that since Q_N and Q_P are perfectly matched and are operating at equal values of $|V_{GS}| = 2.5$ V, the circuit is symmetrical, which dictates that $V_O = 0$ V. Thus both Q_N and Q_P are operating with $|V_{DG}| = 0$ and, hence, in saturation. The drain currents can now be found from

$$I_{DP} = I_{DN} = \tfrac{1}{2} \times 1 \times (2.5 - 1)^2 = 1.125 \text{ mA}$$

Next, consider the circuit with $V_I = +2.5$ V. Transistor Q_P will have a V_{SG} of zero and thus will be cut off, reducing the circuit to that shown in Fig. 5.26(c). Note that V_O will be negative, and thus V_{GD} will

be greater than V_{tn}, causing Q_N to operate in the triode region. For simplicity we shall assume that V_{DS} is small and thus use

$$I_{DN} \simeq k_n'(W_n/L_n)(V_{GS} - V_{tn})V_{DS}$$
$$= 1[2.5 - (-2.5) - 1][V_O - (-2.5)]$$

From the circuit diagram shown in Fig. 5.26(c), we can also write

$$I_{DN}(\text{mA}) = \frac{(0 - V_O)(\text{V})}{10(\text{k}\Omega)}$$

These two equations can be solved simultaneously to yield

$$I_{DN} = 0.244 \text{ mA} \qquad V_O = -2.44 \text{ V}$$

Note that $V_{DS} = -2.44 - (-2.5) = 0.06$ V, which is small as we assumed.

Finally, the situation for the case $V_I = -2.5$ V [Fig. 5.26(d)] will be the exact complement of the case $V_I = +2.5$ V: Transistor Q_N will be off. Thus $I_{DN} = 0$, and Q_P will be operating in the triode region with $I_{DP} = 0.244$ mA and $V_O = +2.44$ V.

Video Example VE 5.4

For the circuits in Fig. VE5.4, $\mu_n C_{ox} = 3\,\mu_p C_{ox} = 270\ \mu\text{A/V}^2$, $|V_t| = 0.5$ V, $\lambda = 0$, $L = 1\ \mu$m, and $W = 3\ \mu$m, unless otherwise specified. Find the labeled currents and voltages.

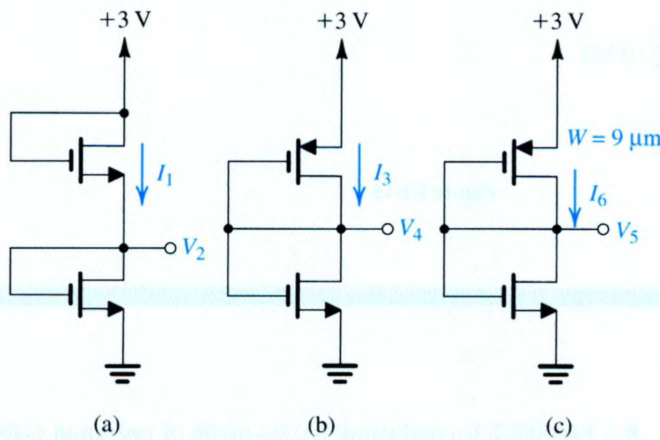


Figure VE5.4 Circuits for Video Example 5.4.

Video Example VE 5.4 *continued*


VE 5.4

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 5.58

EXERCISE

5.15 The NMOS and PMOS transistors in Fig. E5.15 are matched with $k_n'(W_n/L_n) = k_p'(W_p/L_p) = 1\ \text{mA/V}^2$ and $V_{tn} = -V_{tp} = 1$ V. Assuming $\lambda = 0$ for both devices, find the drain currents I_{DN} and I_{DP} and the voltage V_O for $V_I = 0$ V, +2.5 V, and –2.5 V.

Ans. $V_I = 0$ V: 0 mA, 0 mA, 0 V; $V_I = +2.5$ V: 0.104 mA, 0 mA, 1.04 V; $V_I = -2.5$ V: 0 mA, 0.104 mA, –1.04 V

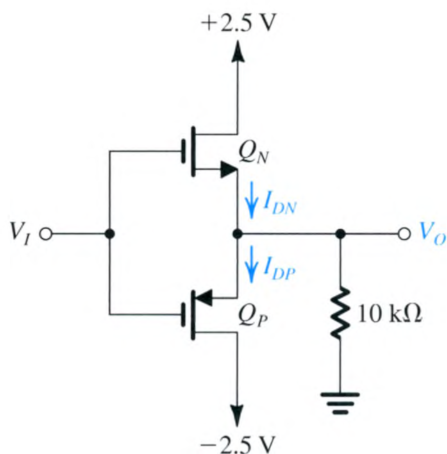


Figure E5.15

Concluding Remark If a MOSFET is conducting but its mode of operation (saturation or triode) is not known, we assume operation in the saturation region, solve the problem, and check whether the conditions for saturation-mode operation are satisfied. If not, then the MOSFET is operating in the triode region, and we redo the analysis accordingly.

GORDON MOORE'S LAW

A half-century ago, Gordon Moore presented a startling idea in the issue of *Electronics Magazine* dated April 19, 1965. Moore, who would become a cofounder first of Fairchild Semiconductor and then of Intel, had projected the potential growth of the integrated-circuit industry based on five points spanning a seven-year period from 1958 to 1965. He concluded the number of transistors per chip had been increasing and would continue to increase by a factor of 2 every two years or so. This finding was destined to propel progress in integrated circuits over the succeeding decades into the twenty-first century. His prediction that the number of transistors would continue to double was based on another prediction: the continuing shrinkage of transistor dimensions. In recognizing the importance of Moore's hypothesis Carver Mead, a pioneer in very-large-scale integration (VLSI), soon began to refer to this prediction as "Moore's law."

5.4 Technology Scaling (Moore's Law) and Other Topics



In this section, we round out our study of the operation and characteristics of the MOSFET by briefly looking into a number of related topics, the most important of which is technology scaling. Popularly known as Moore's law, the continual reduction of the physical dimensions of MOSFETs over the past 50 years has been the driving force of the microelectronics revolution.

5.4.1 Technology Scaling

As mentioned in Section 5.1.4, a given CMOS process technology is characterized by the minimum channel length it allows. Thus, in a 0.18-μm CMOS process, the minimum allowable MOSFET channel length is 0.18 μm. Advances in CMOS fabrication technology over the past 50 years have resulted in processes for which L_{min} has been continually reduced. In fact, every 2 to 3 years, a new CMOS process has been introduced for which L_{min} is reduced by about 30%, that is, to 0.7 of the value in the preceding technology. Thus, with every new **technology generation**, the MOSFET area has been reduced by a factor of $1/(0.7 \times 0.7)$ or approximately 2, allowing the assembly of twice as many devices on a silicon chip of the same area. This astounding phenomenon, predicted over 50 years ago by Gordon Moore, has become known as **Moore's law**. It is the ability to pack an exponentially increasing number of transistors on an integrated-circuit (IC) chip that has resulted in the continual reduction in the cost of electronic products (e.g., computers, tablets, smartphones, etc.). Also, the diminishing size of transistors has enabled these devices to operate at much greater speeds.

Figure 5.27 shows the exponential reduction in MOSFET channel length (by a factor of 2 every 5 years) over a period of 50 years. The solid dots on the graph indicate some of the prominent technology generations or **technology nodes**. Thus, we see the 10-μm process of the early 1970s, the submicron ($L_{min} < 1$ μm) processes of the early 1990s, and the deep-submicron ($L_{min} < 0.25$ μm) processes of the last two decades, including the state-of-the-art 14-nm process represented by the second-last dot on the graph. A number of microprocessor chips manufactured using this technology were announced in 2018. These

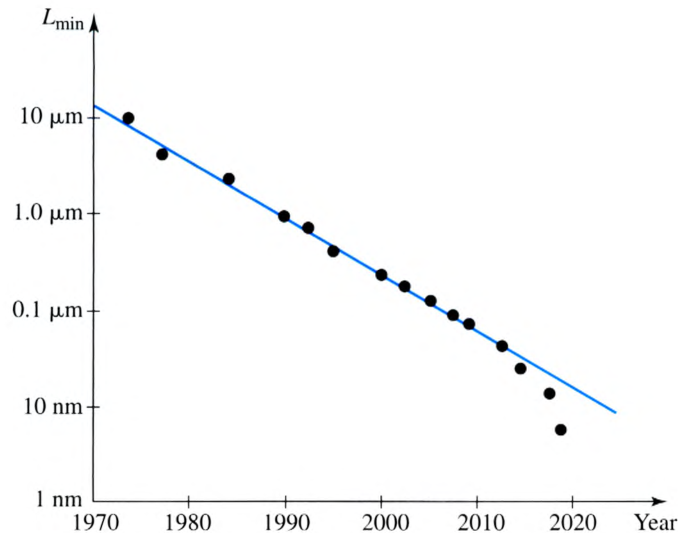


Figure 5.27 The MOSFET channel length has been reduced by a factor of 2 about every 5 years. This phenomenon, known as Moore's law, is continuing.

chips contain between 6 billion and 9 billion transistors and are clocked at about 5 GHz. Finally, the last dot on the graph represents a 7-nm CMOS technology that in 2018 was used to fabricate a computer memory chip. We will study digital circuits in Part III of this book.

Implications of Technology Scaling Technology scaling refers to the scaling down of the MOSFET device dimensions L, W, and t_{ox} by a factor $1/S$, where $S > 1$. Simultaneously, the power supply voltage V_{DD} and the threshold voltage V_t are scaled down by the same factor. The reduction of V_{DD} is motivated by two factors: (1) to keep the longitudinal electric field in the short channel from reaching very high values and (2) to reduce the power dissipation.

Table 5.3 shows some of the scaling implications. Notice that the MOSFET area is scaled by $1/S^2$, and that the oxide capacitance C_{ox} is scaled by the factor S. Assuming μ_n and μ_p do not change, the process transconductance parameters, k_n' and k_p', increase by the factor S. Finally, the gate capacitance C_{gate} is scaled by $1/S$ and thus decreases. The reduction in C_{gate} points in the direction of increased speed of operation, as we will see in later chapters.

Typical Parameter Values Realized in Some CMOS Processes With the theoretical implications of technology scaling in hand, let's now look at some of the more popular CMOS processes. Table 5.4 provides typical values for the important parameters of NMOS and PMOS transistors fabricated in a number of CMOS processes, ranging from the rather old 0.8-μm process to the very recent 28-nm process. Although the 0.8-μm and 0.5-μm processes are now obsolete, their data are included to show trends in the values of various parameters.

Table 5.3 Implications of Device and Voltage Scaling

	Parameter	Relationship	Scaling Factor
1	W, L, t_{ox}		$1/S$
2	V_{DD}, V_t		$1/S$
3	Area/Device	WL	$1/S^2$
4	C_{ox}	ϵ_{ox}/t_{ox}	S
5	k'_n, k'_p	$\mu_n C_{ox}, \mu_p C_{ox}$	S
6	C_{gate}	WLC_{ox}	$1/S$

Table 5.4 Typical Values of CMOS Device Parameters

	0.8 μm		0.5 μm		0.25 μm		0.18 μm		0.13 μm		65 nm		28 nm	
Parameter	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS
t_{ox} (nm)	15	15	9	9	6	6	4	4	2.7	2.7	1.4	1.4	1.4	1.4
C_{ox} (fF/μm²)	2.3	2.3	3.8	3.8	5.8	5.8	8.6	8.6	12.8	12.8	25	25	34	34
μ (cm²/V·s)	550	250	500	180	460	160	450	100	400	100	216	40	220	200
μC_{ox} (μA/V²)	127	58	190	68	267	93	387	86	511	128	540	100	750	680
V_{t0} (V)	0.7	−0.7	0.7	−0.8	0.5	−0.6	0.5	−0.5	0.4	−0.4	0.35	−0.35	0.3	−0.3
V_{DD} (V)	5	5	3.3	3.3	2.5	2.5	1.8	1.8	1.3	1.3	1.0	1.0	0.9	0.9
$\lvert V'_A \rvert$ (V/μm)	25	20	20	10	5	6	5	6	5	6	3	3	1.5	1.5
C_{ov} (fF/μm)	0.2	0.2	0.4	0.4	0.3	0.3	0.37	0.33	0.36	0.33	0.33	0.31	0.4	0.4

Here we note that although the 0.18-μm and 0.13-μm processes are not very new, they remain popular in the design of analog circuits.[10]

Some important observations on the data in Table 5.4 are:

1. The scaling down of t_{ox} and the corresponding scaling up of C_{ox} follows theoretical predictions, except for the last technology node shown (28-nm), which represents the trend for subsequent technologies: t_{ox} cannot be reduced further as this results in increased gate leakage current. Thus, to increase C_{ox}, a different material (other than silicon dioxide) with a higher permittivity is used for the gate dielectric.
2. Although not predicted by our simple scaling framework (Table 5.3), both μ_n and μ_p decrease with scaling, with the latter decreasing much faster. Thus, we see that for the 0.13-μm and 65-nm processes, the μ_p/μ_n ratio is nearly $1/4$. This is problematic as it

[10]An important caution is in order regarding the data in Table 5.4: These data *do not* pertain to any particular commercially available process. Accordingly, these generic data are not intended for use in an actual IC design; rather, they show trends and, as we shall see, help to illustrate design trade-offs while enabling us to work out design examples and problems with parameter values that are as realistic as possible.

results in PMOS devices becoming the bottleneck for the gain and speed achieved in CMOS circuits (a lot more on this later). To reverse this trend, modern processes (including the 28-nm process shown) use techniques that result in μ_p becoming almost equal to μ_n.

3. While the reduction in V_{DD} has been dramatic (from 5 V for the 0.8-μm process to 0.9 V for the 28-nm process), it has not followed the theoretical trend assumed in Table 5.3. The same observation applies to $|V_t|$, which despite having been reduced from 0.7 V to 0.3 V, has not followed the theoretical scaling model. There is a very good reason for not reducing $|V_t|$ any further: as $|V_t|$ becomes smaller, the leakage current through the channel becomes greater. We will have more to say about leakage currents shortly.
4. Another significant though undesirable feature of modern deep-submicron CMOS technologies is that the channel-length modulation effect has become very pronounced. As a result, $|V_A'|$ has decreased to about 1.5 V/μm. Correspondingly, short-channel MOSFETs exhibit low values for the output resistance r_o.

The $i_D - v_{GS}$ Characteristic of Short-Channel MOSFETs In Section 5.1 we arrived at the square-law $i_D - v_{GS}$ characteristic by assuming that the drift velocity of the electrons along the channel, and hence i_D, increases in proportion to the longitudinal electric field established by v_{DS}. However, this does not apply for very short channels; at high values of the longitudinal electric field, the carrier drift velocity no longer increases. This **velocity saturation** phenomenon is one of a host of issues that occur in short-channel devices, resulting in i_D increasing more slowly with v_{GS} than predicted by the square-law relationship.

It is beyond the scope of this book to consider more elaborate models that better characterize short-channel MOSFETs. Thus, for pencil-and-paper analysis, which is critical in the initial phase of designing a circuit, we will continue to use the square-law MOSFET model. For circuit simulation, which is an indispensable step in the final stages of designing an IC, more elaborate models available in simulation programs such as Spice can be used (see Appendix B).

5.4.2 Subthreshold Conduction and Leakage Currents

In our study of the NMOS transistor in Section 5.1, we assumed that current conduction between drain and source occurs only when v_{GS} exceeds V_t. That is, we assumed that for $v_{GS} < V_t$ no current flows between drain and source. This, however, turns out not to be the case, especially for deep-submicron devices. Specifically, for $v_{GS} < V_t$, a small current i_D flows. To be able to see this **subthreshold conduction**, we have redrawn the $i_D - v_{GS}$ graph in Fig. 5.28 using a logarithmic scale for i_D. Notice that at low values of v_{GS}, the relationship between log i_D and v_{GS} is linear, indicating that i_D varies exponentially with v_{GS}. The exponential $i_D - v_{GS}$ characteristic of a MOSFET operating in the subthreshold region resembles that of the BJT (Chapter 6).

Subthreshold conduction has been put to good use in the design of very-low-power circuits such as those needed for electronic watches. Generally speaking, however, subthreshold conduction is a problem in digital IC design, primarily because of the nonzero current i_D that flows for $v_{GS} = 0$. This current is considered a leakage current and results in power dissipation even when the transistor is cut off. To keep leakage currents as small as possible, $|V_t|$ of the MOSFET is not reduced below 0.3 V or so. Although the leakage current is low (few pico amps) and the power dissipation in each cutoff transistor is correspondingly

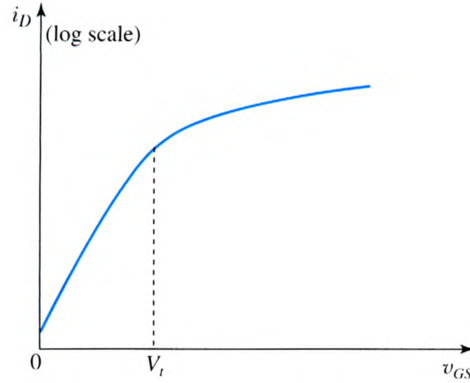


Figure 5.28 The i_D–v_{GS} characteristic of a short-channel MOSFET. To show the details of subthreshold conduction a logarithmic scale is needed for i_D.

small, the problem becomes serious in modern digital IC chips, which contain billions of transistors. Leakage currents, therefore, limit the battery life of mobile electronics that are idle for long periods of time. Also, leakage currents increase dramatically with temperature (similar to the diode saturation or scale current), so that more power is wasted at higher temperatures.

5.4.3 The Role of the Substrate—The Body Effect

In many applications the source terminal of a MOSFET is connected to the substrate (or body) terminal B, which causes the *pn* junction between the substrate and the induced channel (review Fig. 5.5) to have a constant zero (cutoff) bias. In such a case the substrate does not play any role in circuit operation and its existence can be ignored altogether.

In integrated circuits, however, the substrate is usually shared across many MOS transistors. In order to maintain the cutoff condition for all the substrate-to-channel junctions, the substrate is usually connected to the most negative power supply in an NMOS circuit (the most positive in a PMOS circuit). The resulting reverse-bias voltage between source and body (V_{SB} in an *n*-channel device) will have an effect on device operation. Consider, for example, an NMOS transistor whose substrate is negative relative to its source. The reverse-bias voltage will widen the depletion region (refer to Fig. 5.2). This, in turn, reduces the channel depth. To return the channel to its former state, v_{GS} has to be increased.

The simplest way to represent the effect of V_{SB} on the channel is as a change in the threshold voltage V_t. Increasing the reverse substrate bias voltage V_{SB} results in an increase in V_t according to the relationship

$$V_t = V_{t0} + \gamma\left[\sqrt{2\phi_f + V_{SB}} - \sqrt{2\phi_f}\right] \tag{5.30}$$

where V_{t0} is the threshold voltage for $V_{SB} = 0$; ϕ_f is a physical parameter with $(2\phi_f)$ typically 0.6 V; γ is a fabrication-process parameter given by

$$\gamma = \frac{\sqrt{2qN_A\epsilon_s}}{C_{ox}} \tag{5.31}$$

where q is the magnitude of the electron charge (1.6×10^{-19} C), N_A is the doping concentration of the *p*-type substrate, and ϵ_s is the permittivity of silicon ($11.7\epsilon_0 = 11.7$

$\times$ 8.854 $\times$ 10^{-14} = 1.04 $\times$ 10^{-12} F/cm). The parameter γ has the dimension of $\sqrt{\text{V}}$ and is typically 0.4 $\text{V}^{1/2}$. Finally, note that Eq. (5.30) applies equally well for *p*-channel devices with V_{SB} replaced by the reverse bias of the substrate, V_{BS} (or, alternatively, replace V_{SB} by $|V_{SB}|$) and note that γ is negative. Also, in evaluating γ, N_A must be replaced with N_D, the doping concentration of the *n* well in which the PMOS is formed. For *p*-channel devices, $2\phi_f$ is typically 0.75 V, and γ is typically –0.5 $\text{V}^{1/2}$.

EXERCISE

5.16 An NMOS transistor has $V_{t0} = 0.8$ V, $2\phi_f = 0.7$ V, and $\gamma = 0.4\ \text{V}^{1/2}$. Find V_t when $V_{SB} = 3$ V.
Ans. 1.23 V

Equation (5.30) indicates that an incremental change in V_{SB} causes an incremental change in V_t, which in turn produces an incremental change in i_D, even though v_{GS} might have been kept constant. It follows that the body voltage controls i_D; thus, the body acts as another gate for the MOSFET, a phenomenon known as the **body effect**. The parameter γ is known as the **body-effect parameter**.

5.4.4 Temperature Effects

Both V_t and k' are temperature sensitive. The magnitude of V_t decreases by about 2 mV for every 1°C rise in temperature. This decrease in $|V_t|$ gives rise to a corresponding increase in drain current as temperature is increased. However, because k' decreases with temperature and its effect is a dominant one, the overall observed effect of a temperature increase is a *decrease* in drain current. This very interesting result is put to use in applying the MOSFET in power circuits (Chapter 12).

5.4.5 Breakdown and Input Protection

As the voltage on the drain is increased, a value is reached at which the *pn* junction between the drain region and the substrate suffers avalanche breakdown (see Section 3.5.3). This breakdown usually occurs at voltages of 10 V and higher and results in a somewhat rapid increase in current (known as a **weak avalanche**).

Another breakdown effect that occurs at lower voltages in modern devices is called **punch-through**. It occurs in devices with relatively short channels when the drain voltage is increased to the point that the depletion region surrounding the drain region extends through the channel to the source. The drain current then increases rapidly. Normally, punch-through does not result in permanent damage to the device.

Yet another kind of breakdown occurs when the gate-to-source voltage exceeds a few volts. This is the breakdown of the gate oxide and results in permanent damage to the device. As the MOSFET has a very high input resistance and a very small input capacitance, small amounts of static charge accumulating on the gate capacitor can cause its breakdown voltage to be exceeded.

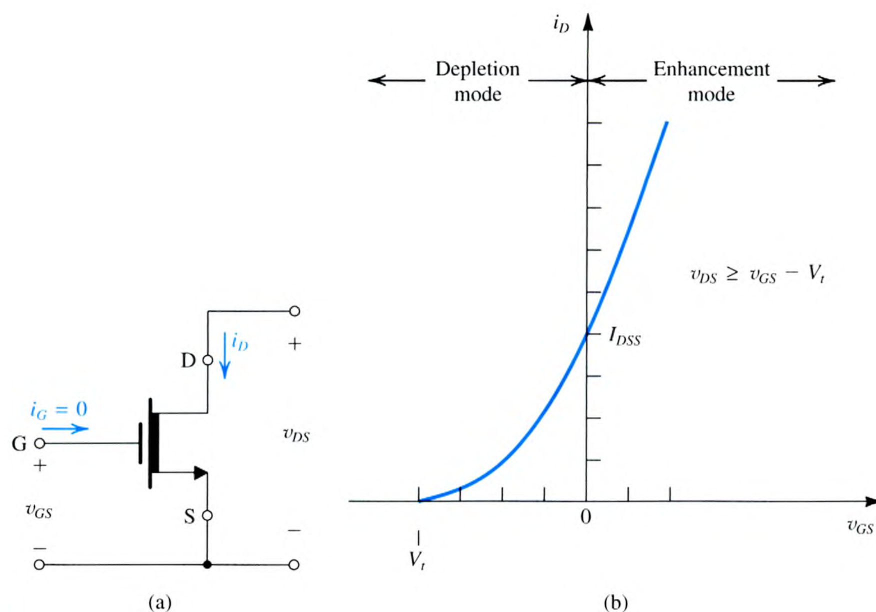


Figure 5.29 The circuit symbol **(a)** and the i_D–v_{GS} characteristic in saturation **(b)** for an n-channel depletion-type MOSFET. Note that I_{DSS} denotes the drain current when $v_{GS} = 0$.

To prevent the accumulation of static charge on the gate capacitor of a MOSFET, gate-protection devices are usually included at the input terminals of MOS integrated circuits. The protection mechanism invariably makes use of clamping diodes.

In later chapters, we shall see that small MOSFET devices are required for high-speed operation, but due to breakdown, it is difficult to combine high speed and high voltages in MOSFET circuits. This is where the bipolar junction transistor (studied in Chapter 6) excels.

5.4.6 The Depletion-Type MOSFET

We conclude this section with a brief discussion of another type of MOSFET, the depletion-type MOSFET. Its structure is similar to that of the enhancement-type MOSFET with one important difference: The depletion MOSFET has a physically implanted channel. Thus an n-channel depletion-type MOSFET has an n-type silicon region connecting the n^+ source and the n^+ drain regions at the top of the p-type substrate. Thus if a voltage v_{DS} is applied between drain and source, a current i_D flows for $v_{GS} = 0$. In other words, there is no need to induce a channel, unlike the case of the enhancement-type MOSFET.

The channel depth and hence its conductivity can be controlled by v_{GS} in exactly the same manner as in the enhancement-type device. Applying a positive v_{GS} enhances the channel by attracting more electrons into it. Here, however, we also can apply a negative v_{GS}, which causes electrons to be repelled from the channel; thus the channel becomes shallower, and its conductivity decreases. The negative v_{GS} is said to **deplete** the channel of its charge carriers, and this mode of operation (negative v_{GS}) is called **depletion mode**. As the magnitude of v_{GS} is increased in the negative direction, a value is reached at which the channel is completely depleted of charge carriers and i_D is reduced to zero, even though v_{DS} may be still applied. This negative value of v_{GS} is the threshold voltage of the n-channel depletion-type MOSFET.

The description above suggests (correctly) that a depletion-type MOSFET can be operated in the enhancement mode by applying a positive v_{GS} and in the depletion mode

by applying a negative v_{GS}. This is illustrated in Fig. 5.29, which shows both the circuit symbol for the depletion NMOS transistor (Fig. 5.29a) and its i_D–v_{GS} characteristic. Observe that here the threshold voltage V_{tn} is negative. The i_D–v_{DS} characteristics (not shown) are similar to those for the enhancement-type MOSFET except for the negative V_{tn}. Finally, note that in the device symbol the shaded area next to the vertical line denotes the existing channel.

Depletion-type MOSFETs can be fabricated on the same IC chip as enhancement-type devices, resulting in circuits with improved characteristics. The depletion-type MOSFET, however, is a specialty device and is not commonly used.

EXERCISE

5.17 For a depletion-type NMOS transistor with $V_t = -2$ V and $k'_n(W/L) = 2$ mA/V^2, find the minimum v_{DS} required to operate in the saturation region when $v_{GS} = +1$ V. What is the corresponding value of i_D?
Ans. 3 V; 9 mA

Summary

- The enhancement-type MOSFET is the most widely used semiconductor device. It is the basis of CMOS technology, which is the most popular IC fabrication technology today. CMOS provides both *n*-channel (NMOS) and *p*-channel (PMOS) transistors, which increases design flexibility. The minimum MOSFET channel length achievable with a given CMOS process is used to characterize the process. This figure has been continually reduced and is currently 14 nm.

- The overdrive voltage, $|v_{OV}| \equiv |v_{GS}| - |V_t|$, is the key quantity that governs the operation of the MOSFET. For the MOSFET to operate in the saturation region, which is the region for amplifier application, $|v_{DS}| \geq |v_{OV}|$, and the resulting $i_D = \frac{1}{2}\mu_n C_{ox}(W/L)v_{OV}^2$ (for NMOS; replace μ_n with μ_p for PMOS). If $|v_{DS}| < |v_{OV}|$, the MOSFET operates in the triode region, which together with cutoff is used for operating the MOSFET as a switch.

- Tables 5.1 and 5.2 summarize the conditions and relationships that describe the operation of NMOS and PMOS transistors, respectively.

- In saturation, i_D shows some linear dependence on v_{DS} as a result of the change in channel length. This channel-length modulation becomes more pronounced as L decreases. It is modeled by ascribing an output resistance $r_o = |V_A|/I_D$ to the MOSFET model. Here, the Early voltage $|V_A| = |V'_A|L$, where $|V'_A|$ is a process-dependent parameter.

- When analyzing dc MOSFET circuits, if a MOSFET is conducting but its region of operation (saturation or triode) is not known, we assume saturation-mode operation. Then, we solve the problem and check to determine whether the assumption was justified. If not, then the transistor is operating in the triode region, and we must redo the analysis accordingly.

- Over the past 50 years, the MOSFET dimensions have been continually reduced, allowing IC designers to pack more and more transistors (now, billions) on the same chip and to operate the resulting chips at vastly increased speeds. Refer to Table 5.3 for the implications of technology scaling and to Table 5.4 for representative parameter values of MOSFETs fabricated in a number of popular technology nodes.

- The depletion-type MOSFET has an implanted channel and thus can be operated in either the depletion or enhancement mode. It is characterized by the same equations used for the enhancement device except for having a negative V_{tn} (positive V_{tp} for depletion PMOS transistors).

Computer Simulations Problems

SIM Problems identified by this icon are intended to demonstrate the value of using SPICE simulation to verify hand analysis and design, and to investigate important issues such as allowable signal swing and amplifier nonlinear distortion. Instructions to assist in setting up simulations for all the indicated problems can be found in the corresponding files on the website. Note that if a particular parameter value is not specified in the problem statement, you are to make a reasonable assumption.

Section 5.1: Device Structure and Physical Operation

5.1 MOS technology is used to fabricate a capacitor, utilizing the gate metallization and the substrate as the capacitor electrodes. Find the area required per 1-pF capacitance for oxide thickness ranging from 1 nm to 10 nm. For a square plate capacitor of 10 pF, what dimensions are needed?

5.2 Calculate the total charge stored in the channel of an NMOS transistor having $C_{ox} = 25\ \text{fF}/\mu\text{m}^2$, $L = 65$ nm, and $W = 650$ nm, and operated at $v_{OV} = 0.15$ V and $v_{DS} = 0$ V.

5.3 Use dimensional analysis to show that the units of the process transconductance parameter k_n' are A/V^2. What are the dimensions of the MOSFET transconductance parameter k_n?

5.4 An NMOS transistor that is operated with a small v_{DS} is found to exhibit a resistance r_{DS}. By what factor will r_{DS} change in each of the following situations?

(a) v_{OV} is doubled.
(b) The device is replaced with another fabricated in the same technology but with double the width.
(c) The device is replaced with another fabricated in the same technology but with both the width and length doubled.
(d) The device is replaced with another fabricated in a more advanced technology for which the oxide thickness is halved and similarly for W and L (assume μ_n remains unchanged).

D 5.5 An NMOS transistor fabricated in a technology for which $k_n' = 511\ \mu\text{A/V}^2$ and $V_t = 0.4$ V is required to operate with a small v_{DS} as a variable resistor ranging in value from 250 Ω to 1 kΩ. Specify the range required for the control voltage v_{GS} and the required transistor width W. It is required to use the smallest possible device, as limited by the minimum channel length of this technology ($L_{\min} = 0.13\ \mu\text{m}$) and the maximum allowed voltage of 1.3 V.

5.6 Sketch a set of i_D-v_{DS} characteristic curves for an NMOS transistor operating with a small v_{DS} (in the manner shown in Fig. 5.4). Let the MOSFET have $k_n = 10\ \text{mA/V}^2$ and $V_{tn} = 0.4$ V. Sketch and clearly label the graphs for $v_{GS} = 0.4$, 0.6, 0.8, 1.0, and 1.2 V. Let v_{DS} be in the range 0 to 50 mV. Give the value of r_{DS} obtained for each of the five values of v_{GS}. Although only a sketch, your diagram should be drawn to scale as much as possible.

D5.7 An n-channel MOS device in a technology for which oxide thickness is 1.4 nm, minimum channel length is 65 nm, $k_n' = 540\ \mu\text{A/V}^2$, and $V_t = 0.35$ V operates in the triode region, with small v_{DS} and with the gate–source voltage in the range 0 V to +1.0 V. If the device has the minimum channel length, what must its width be so that the minimum available resistance is 100 Ω?

5.8 Consider an NMOS transistor operating in the triode region with a constant overdrive voltage V_{OV}. Find an expression for the incremental resistance

$$r_{ds} \equiv 1 \Big/ \frac{\partial i_D}{\partial v_{DS}} \bigg|_{v_{DS}=V_{DS}}$$

Give the values of r_{ds} in terms of k_n and V_{OV} for $V_{DS} = 0$, $0.2V_{OV}$, $0.5V_{OV}$, $0.8V_{OV}$, and V_{OV}.

5.9 An NMOS transistor with $k_n = 5\ \text{mA/V}^2$ and $V_t = 0.4$ V is operated with $v_{GS} = 0.6$ V. At what value of v_{DS} does the transistor enter the saturation region? What value of i_D is obtained in saturation?

5.10 Consider a CMOS process for which $L_{\min} = 0.18\ \mu\text{m}$, $t_{ox} = 4$ nm, $\mu_n = 450\ \text{cm}^2/\text{V}\cdot\text{s}$, and $V_t = 0.5$ V.

(a) Find C_{ox} and k_n'.
(b) For an NMOS transistor with $W/L = 2.4\ \mu\text{m}/0.18\ \mu\text{m}$, calculate the values of v_{OV}, v_{GS}, and $v_{DS\min}$ needed to operate the transistor in the saturation region with a current $i_D = 0.1$ mA.
(c) For the device in (b), find the values of v_{OV} and v_{GS} required to cause the device to operate as a 500-Ω resistor for very small v_{DS}.

---

problems with blue numbers are considered essential; * = difficult problem; ** = more difficult; *** = very challenging
SIM = simulation; **D** = design problem; ● = see related video example

5.11 A p-channel MOSFET with a threshold voltage $V_{tp} = -0.5$ V has its source connected to ground.

(a) What should the gate voltage be for the device to operate with an overdrive voltage of $|v_{OV}| = 0.4$ V?
(b) With the gate voltage as in (a), what is the highest voltage allowed at the drain while the device operates in the saturation region?
(c) If the drain current obtained in (b) is 0.2 mA, what would the current be for $v_D = -20$ mV and for $v_D = -2$V?

5.12 With the knowledge that $\mu_p = 0.25\mu_n$, what must be the relative width of n-channel and p-channel devices having equal channel lengths if they are to have equal drain currents when operated in the saturation mode with overdrive voltages of the same magnitude?

5.13 An n-channel device has $k_n' = 400\ \mu\text{A/V}^2$, $V_t = 0.5$ V, and $W/L = 20$. The device is to operate as a switch for small v_{DS}, utilizing a control voltage v_{GS} in the range 0 V to 1.8 V. Find the switch closure resistance, r_{DS}, and closure voltage, v_{DS}, obtained when $v_{GS} = 1.8$ V and $i_D = 0.2$ mA. If $\mu_p \simeq 0.25\mu_n$, what must W/L be for a p-channel device that provides the same performance as the n-channel device in this application?

5.14 Consider an n-channel MOSFET with $t_{ox} = 4$ nm, $\mu_n = 450\ \text{cm}^2/\text{V}\cdot\text{s}$, $V_t = 0.5$ V, and $W/L = 10$. Find the drain current in the following cases:

(a) $v_{GS} = 1.8$ V and $v_{DS} = 1$ V
(b) $v_{GS} = 0.7$ V and $v_{DS} = 1.5$ V
(c) $v_{GS} = 1.8$ V and $v_{DS} = 0.1$ V
(d) $v_{GS} = v_{DS} = 1.8$ V

Section 5.2: Current–Voltage Characteristics

In the following problems, when λ is not specified, assume it is zero.

***5.15** Show that when channel-length modulation is neglected (i.e., $\lambda = 0$), plotting i_D/k_n versus v_{DS} for various values of v_{OV}, and plotting i_D/k_n versus v_{OV} for $v_{DS} \geq v_{OV}$, results in universal representation of the i_D–v_{DS} and i_D–v_{GS} characteristics of the NMOS transistor. That is, the resulting graphs are both technology and device independent. Furthermore, these graphs apply equally well to the PMOS transistor by a simple relabeling of variables. (How?) What is the slope at $v_{DS} = 0$ of each of the i_D/k_n versus v_{DS} graphs? For the i_D/k_n versus v_{OV} graph, find the slope at a point $v_{OV} = V_{OV}$.

5.16 An NMOS transistor having $V_t = 0.5$ V is operated in the triode region with v_{DS} small. With $v_{GS} = 0.7$ V, it is found to have a resistance r_{DS} of 1 kΩ. What value of v_{GS} is required to obtain $r_{DS} = 200\ \Omega$? Find the corresponding resistance values obtained with a device having twice the value of W.

5.17 A particular MOSFET for which $V_{tn} = 0.4$ V and $k_n'(W/L) = 2\ \text{mA/V}^2$ is to be operated in the saturation region. If i_D is to be 50 µA, find the required v_{GS} and the minimum required v_{DS}. Repeat for $i_D = 200$ µA.

5.18 A particular n-channel MOSFET is measured to have a drain current of 360 µA at $v_{GS} = v_{DS} = 1$ V and of 160 µA at $v_{GS} = v_{DS} = 0.8$ V. What are the values of k_n and V_t for this device?

D5.19 For a particular IC-fabrication process, the transconductance parameter $k_n' = 400\ \mu\text{A/V}^2$, and $V_t = 0.5$ V. In an application in which $v_{GS} = v_{DS} = V_{\text{supply}} = 1.8$ V, a drain current of 2 mA is required of a device of minimum length of 0.18 µm. What value of channel width must the design use?

5.20 An NMOS transistor, operating in the linear-resistance region with $v_{DS} = 50$ mV, is found to conduct 50 µA for $v_{GS} = 0.8$ V and 100 µA for $v_{GS} = 1.2$ V. What is the apparent value of threshold voltage V_t? If $k_n' = 500\ \mu\text{A/V}^2$, what is the device W/L ratio? What current would you expect to flow with $v_{GS} = 1$ V and $v_{DS} = 0.2$ V? If the device is operated at $v_{GS} = 1$ V, at what value of v_{DS} will the drain end of the MOSFET channel just reach pinch-off, and what is the corresponding drain current?

5.21 For an NMOS transistor, for which $V_t = 0.5$ V, operating with v_{GS} in the range of 1.0 V to 1.8 V, what is the largest value of v_{DS} for which the channel remains continuous?

5.22 An NMOS transistor, fabricated with $W = 20$ µm and $L = 1$ µm in a technology for which $k_n' = 100\ \mu\text{A/V}^2$ and $V_t = 0.8$ V, is to be operated at very low values of v_{DS} as a linear resistor. For v_{GS} varying from 1.0 V to 4.8 V, what range of resistor values can be obtained? What is the available range if

(a) the device width is halved?
(b) the device length is halved?
(c) both the width and length are halved?

5.23 When the drain and gate of a MOSFET are connected together, a two-terminal device known as a "diode-connected transistor" results. Figure P5.23 shows such devices obtained from MOS transistors of both polarities. Show that

(a) the i–v relationship is given by

$$i = \frac{1}{2}k'\frac{W}{L}(v - |V_t|)^2$$

(b) the incremental resistance r for a device operating at $v = |V_t| + V_{OV}$ is given by

$$r \equiv 1\Big/\left[\frac{\partial i}{\partial v}\right] = 1\Big/\left(k'\frac{W}{L}V_{OV}\right)$$

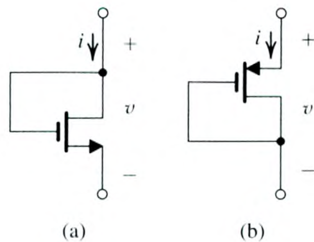


Figure P5.23

5.24 For the circuit in Fig. P5.24, sketch i_D versus v_S for v_S varying from 0 to V_{DD}. Clearly label your sketch.

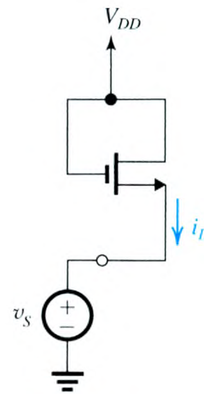


Figure P5.24

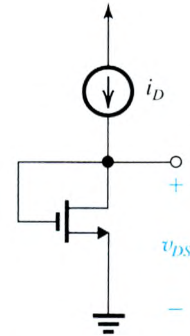


Figure P5.25

5.25 For the circuit in Fig. P5.25, find an expression for v_{DS} in terms of i_D. Sketch and clearly label a graph for v_{DS} versus i_D.

***5.26** The table below lists 10 different cases labeled (a) to (j) for operating an NMOS transistor with $V_t = 1$ V. In each case the voltages at the source, gate, and drain (relative to the circuit ground) are specified. You are required to complete the table entries. Note that if you encounter a case for which v_{DS} is negative, you should exchange the drain and source

	Voltage (V)						
Case	V_S	V_G	V_D	V_{GS}	V_{OV}	V_{DS}	**Region of operation**
a	+1.0	+1.0	+2.0				
b	+1.0	+2.5	+2.0				
c	+1.0	+2.5	+1.5				
d	+1.0	+1.5	0				
e	0	+2.5	+1.0				
f	+1.0	+1.0	+1.0				
g	−1.0	0	0				
h	−1.5	0	0				
i	−1.0	0	+1.0				
j	+0.5	+2.0	+0.5				

before solving the problem. You can do this because the MOSFET is a symmetric device.

5.27 The NMOS transistor in Fig. P5.27 has $V_t = 0.5$ V and $k_n = 2$ mA/V^2. Sketch and clearly label i_D versus v_G with v_G varying in the range 0 to +1.8 V. Give equations for the various portions of the resulting graph.

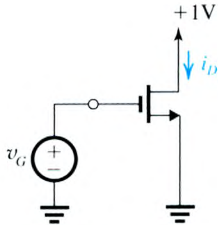


Figure P5.27

5.28 Figure P5.28 shows two NMOS transistors operating in saturation at equal V_{GS} and V_{DS}.

(a) If the two devices are matched except for a maximum possible mismatch in their *W/L* ratios of 2%, what is the maximum resulting mismatch in the drain currents?

(b) If the two devices are matched except for a maximum possible mismatch in their V_t values of 10 mV, what is the maximum resulting mismatch in the drain currents? Assume that the nominal value of V_t is 0.5 V.

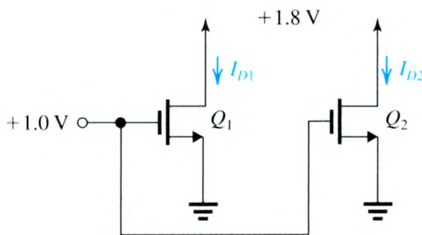


Figure P5.28

5.29 For a particular MOSFET operating in the saturation region at a constant v_{GS}, i_D is found to be 200 μA for $v_{DS} =$ 1 V and 205 μA for $v_{DS} = 1.5$ V. Find the values of r_o, V_A, and λ.

5.30 A particular MOSFET has $V_A = 10$ V. For operation at 10 μA and 100 μA, what are the expected output resistances? In each case, for a change in v_{DS} of 1 V, what percentage change in drain current would you expect?

D5.31 In a particular IC design in which the standard channel length is 0.18 μm, an NMOS device of standard length with *W/L* of 10 operating at 100 μA is found to have an output resistance of 20 kΩ, about $\frac{1}{5}$ of that needed. What dimensional change can be made to solve the problem? What is the new device length? The new device width? The new *W/L* ratio? What is V_A for the standard device in this IC? The new device?

D5.32 For a particular *n*-channel MOS technology, in which the minimum channel length is 0.2 μm, the associated value of λ is 0.5 V^{-1}. If a particular device for which *L* is 1.0 μm operates in saturation at $v_{DS} = 1$ V with a drain current of 100 μA, what does the drain current become if v_{DS} is raised to 2 V? What percentage change does this represent? What can be done to reduce the percentage by a factor of 2?

5.33 An NMOS transistor is fabricated in a 0.18-μm process having $k_n' = 400$ μA/V^2 and $V_A' = 5$ V/μm of channel length. If $L = 0.54$ μm and $W = 5.4$ μm, find V_A and λ. Find the value of i_D that results when the device is operated with an overdrive voltage of 0.25 V and $v_{DS} = 1$ V. Also, find the value of r_o at this operating point. If v_{DS} is increased by 0.5 V, what is the corresponding change in i_D?

5.34 If in an NMOS transistor, both *W* and *L* are quadrupled and V_{OV} is halved, by what factor does r_o change?

D5.35 Consider the circuit in Fig. P5.28 with both transistors perfectly matched but with the dc voltage at the drain of Q_1 lowered to +1 V. If the two drain currents are to be matched within 5% (i.e., the maximum difference allowed between the two currents is 5%), what is the minimum required value of V_A? If the technology is specified to have $V_A' = 10$ V/μm, what is the minimum channel length the designer must use?

5.36 Complete the missing entries in the following table, which describes characteristics of suitably operating NMOS transistors:

MOS	1	2	3	4
λ (V^{-1})		0.1		
V_A (V)	20			5
I_D (mA)	0.5		0.2	
r_o (kΩ)		25	100	500

	V_S	V_G	V_D	V_{SG}	$\|V_{OV}\|$	V_{SD}	Region of operation
a	+2	+2	0				
b	+2	+1	0				
c	+2	0	0				
d	+2	0	+1				
e	+2	0	+1.5				
f	+2	0	+2				

5.37 A PMOS transistor has $k_p'(W/L) = 100\ \mu\text{A/V}^2$, $V_t = -0.5$ V, and $\lambda = -0.1\ \text{V}^{-1}$. The gate is connected to ground and the source to +1 V. Find the drain current for $v_D = +0.8$ V, +0.6 V, +0.5 V, 0 V, and −1 V.

5.38 A *p*-channel transistor for which $|V_t| = 0.4$ V and $|V_A| = 4$ V operates in saturation with $|v_{GS}| = 1$ V, $|v_{DS}| = 1.6$ V, and $i_D = 2$ mA. Find corresponding signed values for v_{GS}, v_{SG}, v_{DS}, v_{SD}, V_t, V_A, λ, and $k_p'(W/L)$.

5.39 The table above lists the terminal voltages of a PMOS transistor in six cases, labeled a, b, c, d, e, and f. The transistor has $V_{tp} = -1$ V. Complete the table entries.

VE 5.1

5.40 The PMOS transistor in Fig. P5.40 has $V_{tp} = -0.4$ V. As the gate voltage v_G is varied from +2 V to 0 V, the transistor moves through all of its three possible modes of operation. Specify the values of v_G at which the device changes modes of operation.

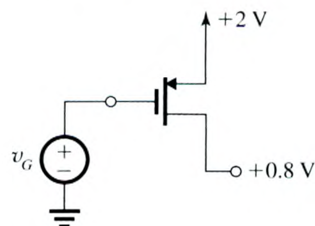


Figure P5.40

****5.41** Various NMOS and PMOS transistors, numbered 1 to 4, are measured in operation, as shown in the table at the top of the next page. For each transistor, find the values of $\mu C_{ox}W/L$ and V_t that apply and complete the table, with V in volts, I in μA, and $\mu C_{ox}W/L$ in $\mu\text{A/V}^2$. Assume $\lambda = 0$.

***5.42** All the transistors in the circuits shown in Fig. P5.42 have the same values of $|V_t|$, k', W/L, and λ. Moreover, λ is negligibly small. All operate in saturation at $I_D = I$ and $|V_{GS}| = |V_{DS}| = 1$ V. Find the voltages V_1, V_2, V_3, and V_4. If $|V_t| = 0.5$ V and $I = 0.1$ mA, how large a resistor can be inserted in series with each drain while maintaining saturation? If the current source I requires at least 0.5 V between its terminals to operate properly, what is the largest resistor that can be placed in series with each MOSFET source while ensuring saturated-mode operation of each transistor at $I_D = I$? In the latter limiting situation, what do V_1, V_2, V_3, and V_4 become?

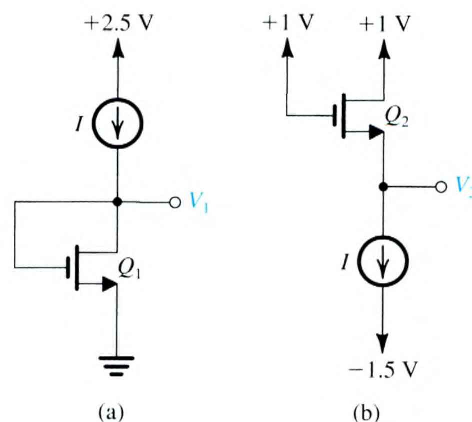


Figure P5.42

Case	Transistor	V_S	V_G	V_D	I_D	Type	Mode	$\mu C_{ox} W/L$	V_t
a	1	0	1	2.5	100				
	1	0	1.5	2.5	400				
b	2	5	3	−4.5	50				
	2	5	2	−0.5	450				
c	3	5	3	4	200				
	3	5	2	0	800				
d	4	−2	0	0	72				
	4	−4	0	−3	270				

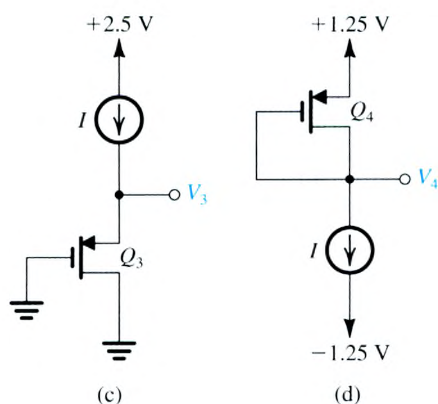


Figure P5.42 *continued*

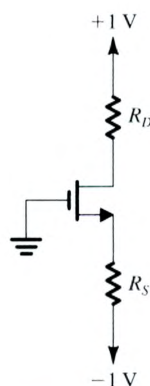


Figure P5.43

Section 5.3: MOSFET Circuits at DC

Note: If λ is not specified, assume it is zero.

D 5.43 Design the circuit of Fig. P5.43 to establish a drain current of 0.1 mA and a drain voltage of +0.2 V. The MOSFET has $V_t = 0.2$ V, $\mu_n C_{ox} = 400\ \mu A/V^2$, $L = 0.5\ \mu m$, and $W = 4\ \mu m$. Specify the required values for R_S and R_D. Assume $\lambda = 0$. VE 5.2

5.44 The NMOS transistor in the circuit of Fig. P5.43 has $V_t = 0.4$ V and $k_n = 4$ mA/V^2. The voltages at the source and the drain are measured and found to be −0.55 V and +0.1 V, respectively. What current I_D is flowing, and what must the values of R_D and R_S be? What is the largest value for R_D for which I_D remains unchanged?

D 5.45 For the circuit in Fig. E5.10, assume that Q_1 and Q_2 are matched except for having different widths, W_1 and W_2. Let $V_t = 0.5$ V, $k_n' = 0.4$ mA/V^2, $L_1 = L_2 = 0.36\ \mu m$, $W_1 = 1.8\ \mu m$, and $\lambda = 0$.

(a) Find the value of R required to establish a current of 100 μA in Q_1.
(b) Find W_2 and R_2 so that Q_2 operates at the edge of saturation with a current of 0.5 mA.

5.46 The transistor in the circuit of Fig. P5.46 has $k_n' = 0.5$ mA/V^2, $V_t = 0.4$ V, and $\lambda = 0$. Show that operation at the edge of saturation is obtained when the following condition is satisfied:

$$\left(\frac{W}{L}\right) R_D = 2.5\ \text{k}\Omega$$

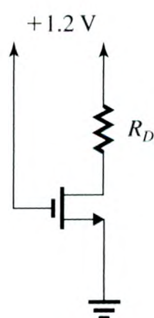


Figure P5.46

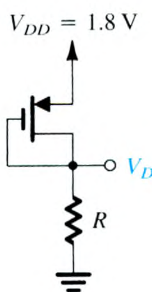


Figure P5.48

D5.47 It is required to operate the transistor in the circuit of Fig. P5.46 at the edge of saturation with $I_D = 50\ \mu A$. If $V_t = 0.4$ V, find the required value of R_D.

D5.48 The PMOS transistor in the circuit of Fig. P5.48 has $V_t = -0.5$ V, $\mu_p C_{ox} = 100\ \mu A/V^2$, $L = 0.18\ \mu m$, and $\lambda = 0$. Find the values required for W and R in order to establish a drain current of 160 μA and a voltage V_D of 0.8 V.

VE 5.3

D5.49 The NMOS transistors in the circuit of Fig. P5.49 have $V_t = 0.4$ V, $\mu_n C_{ox} = 400\ \mu A/V^2$, $\lambda = 0$, and $L_1 = L_2 = 0.18\ \mu m$. Find the required values of gate width for each of Q_1 and Q_2, and the value of R, to obtain the voltage and current values indicated.

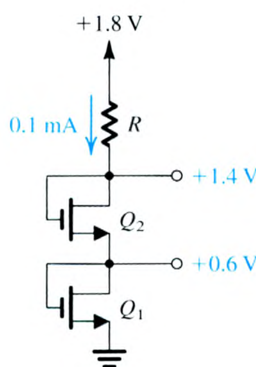


Figure P5.49

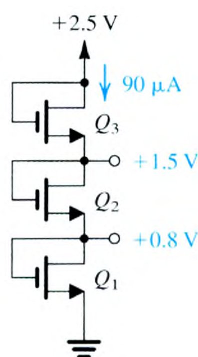


Figure P5.50

D5.50 The NMOS transistors in Fig. P5.50 have $V_t = 0.5$ V, $\mu_n C_{ox} = 200\ \mu A/V^2$, $\lambda = 0$, and $L_1 = L_2 = L_3 = 0.5\ \mu m$. Find the required values of gate width for each of Q_1, Q_2, and Q_3 to obtain the voltage and current values indicated.

5.51 Consider the circuit of Fig. 5.24(a). In Example 5.6 it was found that when $V_t = 1$ V and $k_n'(W/L) = 1\ mA/V^2$, the drain current is 0.5 mA and the drain voltage is +7 V. If the transistor is replaced with another having $V_t = 1.5$ V with $k_n'(W/L) = 1.5\ mA/V^2$, find the new values of I_D and V_D. Comment on how tolerant (or intolerant) the circuit is to changes in device parameters.

D 5.52 Using a PMOS transistor with $V_t = -1.5$ V, $k_p'(W/L) = 4\ mA/V^2$, and $\lambda = 0$, design a circuit that resembles that in Fig. 5.24(a). Using a 10-V supply, design for a gate voltage of +6 V, a drain current of 0.5 mA, and a drain voltage of +5 V. Find the values of R_S and R_D. Also, find the values of the resistances in the voltage divider feeding the gate, assuming a 1-μA current in the divider.

5.53 The MOSFET in Fig. P5.53 has $V_t = 0.4$ V, $k_n' = 500\ \mu A/V^2$, and $\lambda = 0$. Find the required values of W/L and R so that when $V_I = V_{DD} = +1.3$ V, $r_{DS} = 50\ \Omega$ and $V_O = 50$ mV.

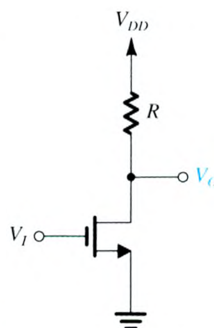


Figure P5.53

***5.54** In the circuits shown in Fig. P5.54, transistors are characterized by $|V_t| = 0.5$ V, $k'W/L = 5\ mA/V^2$, and $\lambda = 0$.

(a) Find the labeled voltages V_1 through V_7.
(b) In each of the circuits, replace the current source with a resistor. Select the resistor value to yield a current as close to that of the current source as possible, while using resistors specified in the 1% table provided in Appendix J.

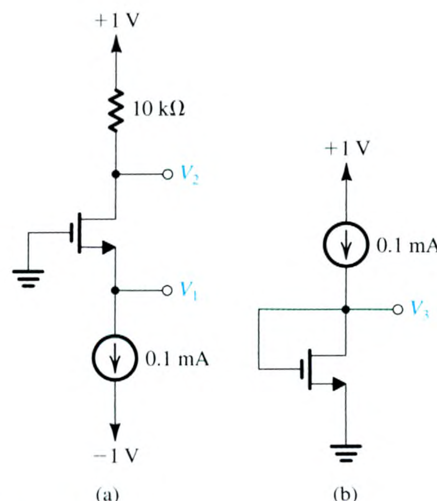


Figure P5.54

CHAPTER 5 PROBLEMS

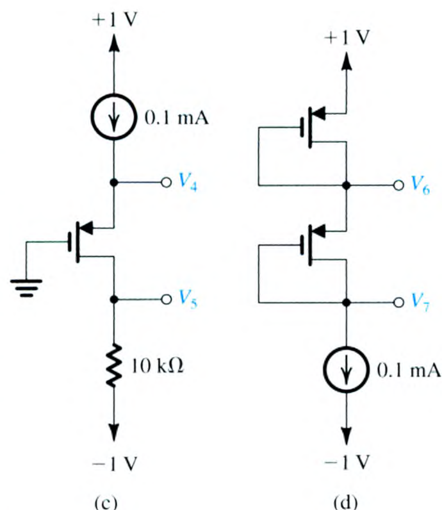


Figure P5.54 *continued*

5.55 For each of the circuits in Fig. P5.55, find the labeled node voltages. For all transistors, $k_n'(W/L) = 2\ \text{mA/V}^2$, $V_t = 0.5\ \text{V}$, and $\lambda = 0$.

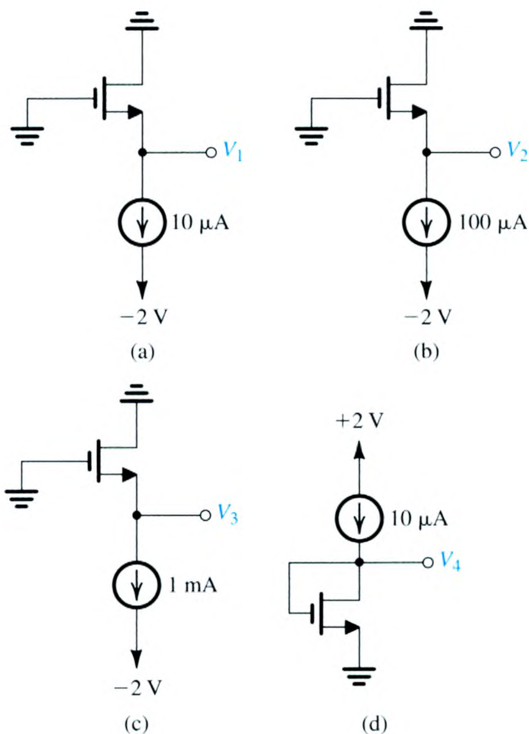


Figure P5.55

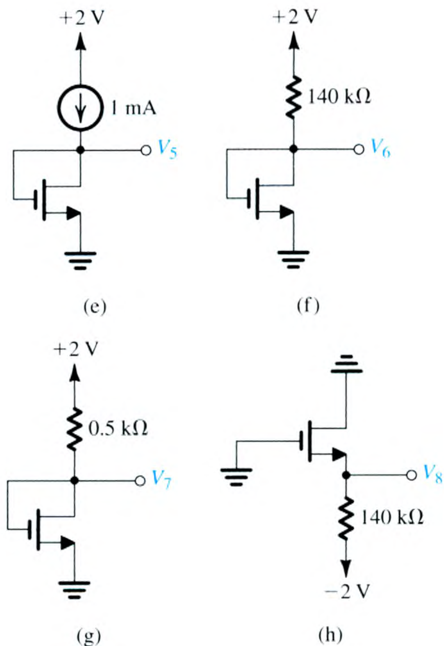


Figure P5.55 *continued*

***5.56** For each of the circuits shown in Fig. P5.56, find the labeled node voltages. The NMOS transistors have $V_t = 0.4\ \text{V}$ and $k_n'(W/L) = 5\ \text{mA/V}^2$.

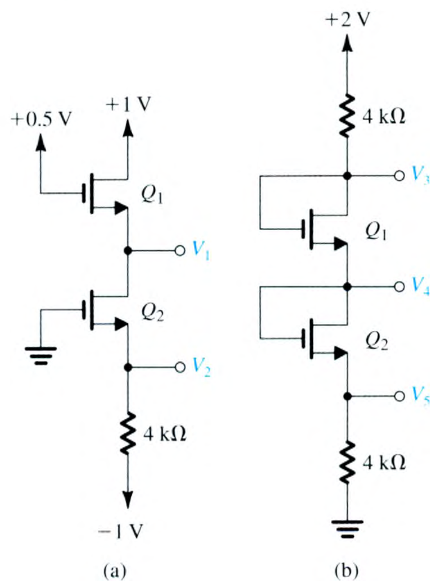


Figure P5.56

***5.57** For the circuit in Fig. P5.57:

(a) Show that for the PMOS transistor to operate in saturation, the following condition must be satisfied:

$$IR \le |V_{tp}|$$

(b) If the transistor is specified to have $|V_{tp}| = 1$ V and $k_p = 0.2$ mA/V^2, and for $I = 0.1$ mA, find the voltages V_{SD} and V_{SG} for $R = 0$, 10 kΩ, 30 kΩ, and 100 kΩ.

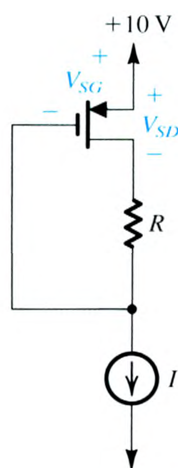


Figure P5.57

5.58 For the circuits in Fig. P5.58, $\mu_n C_{ox} = 4\mu_p C_{ox} = 400\ \mu$A/V^2, $|V_t| = 0.4$ V, $\lambda = 0$, $L = 0.2\ \mu$m, and $W = 1\ \mu$m, unless otherwise specified. Find the labeled currents and voltages.

VE 5.4

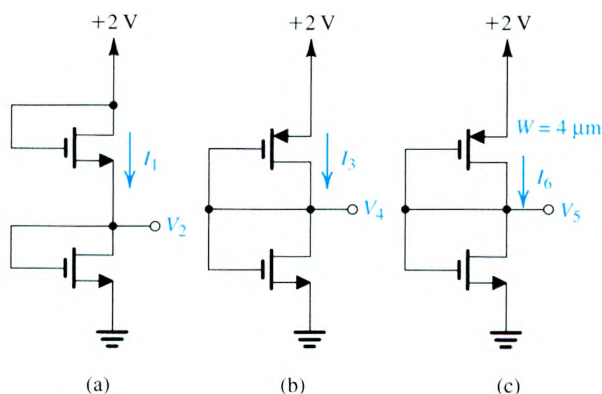


Figure P5.58

SIM ***5.59** For the devices in the circuit of Fig. P5.59, $|V_t| = 0.5$ V, $\lambda = 0$, $\mu_n C_{ox} = 400\ \mu$A/V^2, $L = 0.5\ \mu$m, and $W = 0.5\ \mu$m. Find V_2 and I_2. How do these values change if Q_3 and Q_4 are made to have $W = 5\ \mu$m?

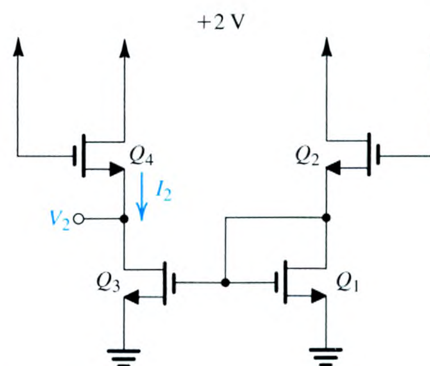


Figure P5.59

5.60 In the circuit of Fig. P5.60, transistors Q_1 and Q_2 have $V_t = 0.5$ V, and the process transconductance parameter $k_n' = 400\ \mu$A/V^2. Find V_1, V_2, and V_3 for each of the following cases:

(a) $(W/L)_1 = (W/L)_2 = 10$
(b) $(W/L)_1 = 1.5(W/L)_2 = 10$

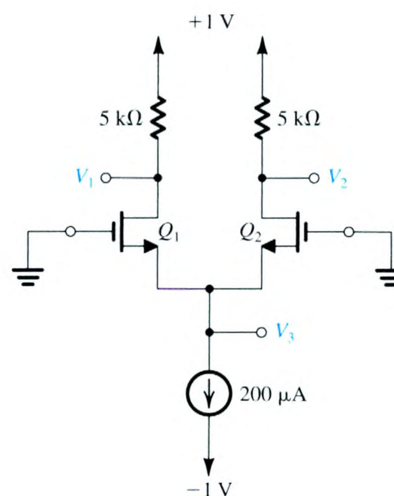


Figure P5.60

CHAPTER 5 PROBLEMS

Section 5.4: Technology Scaling (Moore's Law) and Other Topics

***5.61** The table below shows four technology generations, each characterized by the minimum possible MOSFET channel length (row 1). In going from one generation to another, both L and t_{ox} are scaled by the same factor. The power supply utilized, V_{DD}, is also scaled by the same factor, to keep the magnitudes of all electrical fields within the device unchanged. Unfortunately, but for good reasons, V_t cannot be scaled similarly.

Complete the table entries, noting that in the last row you are asked to find the number of transistors that can be placed on an IC chip fabricated in each of the technologies in terms of the number obtained with the 0.5-μm technology (n).

5.62 A chip with a certain area designed using the 5-μm process of the late 1970s contains 20,000 MOSFETs. What does Moore's law predict the number of transistors to be on a chip of equal area fabricated using the 32-nm process of 2013?

5.63 In a particular application, an n-channel MOSFET operates with V_{SB} in the range 0 V to 4 V. If V_{t0} is nominally 1.0 V, find the range of V_t that results if $\gamma = 0.5\text{ V}^{1/2}$ and $2\phi_f = 0.6$ V. If the gate oxide thickness is increased by a factor of 4, what does the threshold voltage become?

5.64 A p-channel transistor operates in saturation with its source voltage 3 V lower than its substrate. For $\gamma = 0.5\text{ V}^{1/2}$, $2\phi_f = 0.75$ V, and $V_{t0} = -0.7$ V, find V_t.

5.65 Consider a diode-connected NMOS transistor fed with a constant current of 0.5 mA. Assume $\lambda = 0$.

(a) If at 20°C, $V_t = 0.5$ V and $k_n = 1\text{ mA/V}^2$, find V_{GS}.

1	L (μm)	0.5	0.25	0.18	0.13
2	t_{ox} (nm)	10			
3	C_{ox} (fF/μm^2)				
4	k_n' (μA/V^2) ($\mu_n = 500\text{ cm}^2/\text{V}\cdot\text{s}$)				
5	k_n (mA/V^2) For $W/L = 10$				
6	Device area, A (μm^2) $L = L_{\min}$, $W/L = 10$				
7	V_{DD} (V)	5			
8	V_t (V)	0.7	0.5	0.4	0.4
9	I_D (mA) For $V_{GS} = V_{DS} = V_{DD}$				
10	$P\text{ (mW)} = V_{DD}I_D$				
11	P/A (mW/μm^2)				
12	Devices per chip	n			

(b) If the temperature rises to 50°C, find the resulting V_{GS} given that V_t changes by −2mV/°C and k_n' changes by −0.3%/°C.

5.66 (a) Using the expression for i_D in saturation and neglecting the channel-length modulation effect (i.e., let $\lambda = 0$), derive an expression for the per unit change in i_D per °C $\left[(\partial i_D/i_D)/\partial T\right]$ in terms of the per unit change in k_n' per °C $\left[(\partial k_n'/k_n')/\partial T\right]$, the temperature coefficient of V_t in V/°C $(\partial V_t/\partial T)$, and V_{GS} and V_t.
(b) If V_t decreases by 2 mV for every °C rise in temperature, find the temperature coefficient of k_n' that results in i_D decreasing by 0.2%/°C when the NMOS transistor with $V_t =$ 1 V is operated at $V_{GS} = 5$ V.

5.67 A depletion-type n-channel MOSFET with $k_n'W/L =$ 2 mA/V^2 and $V_t = -3$ V has its source and gate grounded. Find the region of operation and the drain current for $v_D = 0.1$ V, 1 V, 3 V, and 5 V. Neglect the channel-length-modulation effect.

5.68 For a particular depletion-mode NMOS device, $V_t =$ −2 V, $k_n'W/L = 200\ \mu$A/V^2, and $\lambda = 0.02$ V^{-1}. When operated at $v_{GS} = 0$, what is the drain current that flows for $v_{DS} = 1$ V, 2 V, 3 V, and 10 V? What does each of these currents become if the device width is doubled with L the same? With L also doubled?

***5.69** Neglecting the channel-length-modulation effect, show that for the depletion-type NMOS transistor of Fig. P5.69 the i–v relationship is given by

$$i = \frac{1}{2}k_n(v^2 - 2V_t v) \qquad \text{for } v \ge V_t$$
$$i = -\frac{1}{2}k_n V_t^2 \qquad \text{for } v \le V_t$$

(Recall that V_t is negative.) Sketch the i–v relationship for the case: $V_t = -2$ V and $k_n = 2$ mA/V^2.

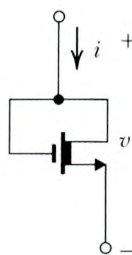


Figure P5.69

CHAPTER 6

Bipolar Junction Transistors (BJTs)

IN THIS CHAPTER YOU WILL LEARN

- The physical structure of the bipolar transistor and how it works.
- How the voltage between two terminals of the transistor controls the current that flows through the third terminal, and the equations that describe these current–voltage characteristics.
- How to analyze and design circuits that contain bipolar transistors, resistors, and dc sources.

Introduction

In this chapter, we study the other major three-terminal device: the bipolar junction transistor (BJT). The material in this chapter parallels but does not rely on the MOSFET material in Chapter 5, so you can study these two chapters in either order.

Three-terminal devices are far more useful than two-terminal ones, such as the diodes we studied in Chapter 4, because they can be used in a multitude of applications, ranging from signal amplification to the design of digital logic and memory circuits. They operate by using the voltage between two terminals to control the current flowing in the third terminal. The three-terminal device can thus be used as a controlled source, which is the basis for amplifier design, as we learned in Chapter 1. Also, in the extreme, we can use the control signal to cause the current in the third terminal to change from zero to a large value, allowing the device to act as a switch. As we will see in Chapter 16, the switch is the basis for the logic inverter, the basic element of digital circuits.

The invention of the BJT in 1948 at the Bell Telephone Laboratories ushered in the era of solid-state circuits. Transistors replaced vacuum tubes in radios and television sets and sparked a revolution in the way we work, play, and, indeed, live. We can trace the dominance of information technology and the emergence of the knowledge-based economy to the invention of the transistor.

The bipolar transistor enjoyed nearly three decades as the device of choice in the design of both discrete and integrated circuits. Although the MOSFET had been known very early on, because its fabrication requires a thin insulating layer, it was not until the 1970s and 1980s that it became a serious competitor to the BJT. By 2019, the MOSFET was undoubtedly the most widely used electronic device, and CMOS technology had become the technology of choice whenever a large number of transistors must be combined on a single chip. Nevertheless, the BJT remains a significant device that excels in certain applications.

The BJT remains popular in discrete-circuit design, where it is used together with other discrete components such as resistors and capacitors to implement circuits that are assembled on printed-circuit boards (PCBs). There is a very wide selection of BJT types suited for nearly every conceivable application. BJTs (on their own and in combination with CMOS in what is known as BiCMOS) remain particularly important for circuits that combine high voltages and high-frequency signals; for example, power amplifiers for radio and higher-frequency applications.

In this chapter, we start with a description of the physical operation of the BJT. This physical description will help you understand the performance of the transistor as a circuit element. We then quickly move from describing current flow in terms of electrons and holes to a study of the transistor terminal characteristics. We will develop circuit models for transistor operation in different modes and use them in the analysis and design of transistor circuits. The main objective of this chapter is to make you very familiar with the BJT and to lay the foundation for the use of the BJT in amplifier design (Chapter 7).

6.1 Device Structure and Physical Operation

6.1.1 Simplified Structure and Modes of Operation

Figure 6.1 shows a simplified structure for the BJT (We will show a practical transistor structure later; see also Appendix A, which deals with fabrication technology.) As you

can see in Fig. 6.1, the BJT consists of three semiconductor regions: the emitter region (*n* type), the base region (*p* type), and the collector region (*n* type). This kind of transistor is called an *npn* transistor. Another transistor, a dual of the *npn*, shown in Fig. 6.2, has a *p*-type emitter, an *n*-type base, and a *p*-type collector, and is appropriately called a *pnp* transistor. A terminal is connected to each of the three semiconductor regions of the transistor, with the terminals labeled **emitter** (E), **base** (B), and **collector** (C).

The transistor consists of two *pn* junctions, called the **emitter–base junction** (EBJ) and the **collector–base junction** (CBJ). The bias condition (forward or reverse) of each of these junctions will give us the different modes of operation of the BJT shown in Table 6.1. We use the **active mode** if we want the transistor to operate as an amplifier. Switching applications (e.g., logic circuits) utilize both the **cutoff mode** and the **saturation mode**. As the name implies, in the cutoff mode no current flows; this is because both junctions are reverse biased.

As we will see shortly, charge carriers of both polarities—that is, electrons and holes—participate in the current-conduction process in a bipolar transistor, which is the reason for the name *bipolar*. (Notice the contrast with the MOSFET, where current is

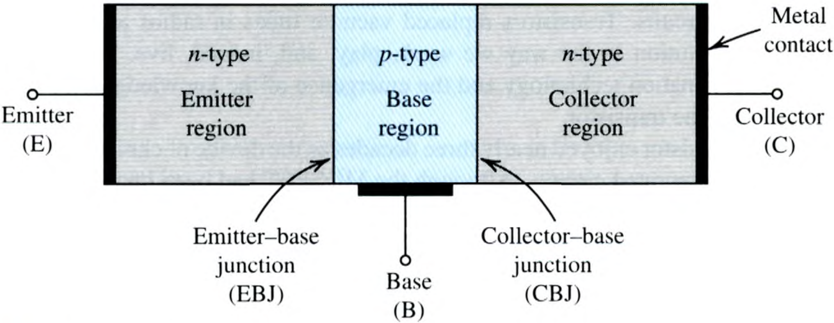


Figure 6.1 A simplified structure of the *npn* transistor.

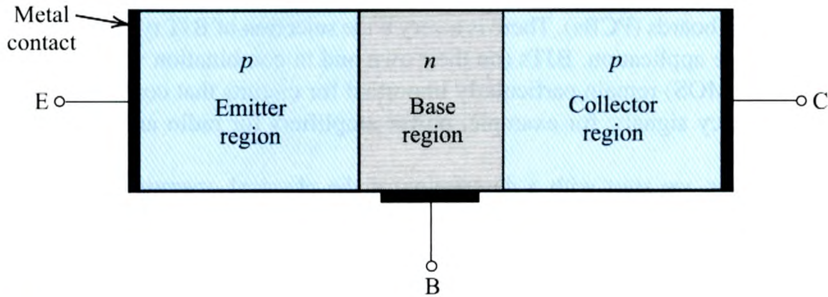


Figure 6.2 A simplified structure of the *pnp* transistor.

Table 6.1 BJT Modes of Operation

Mode	EBJ	CBJ
Cutoff	Reverse	Reverse
Active	Forward	Reverse
Saturation	Forward	Forward

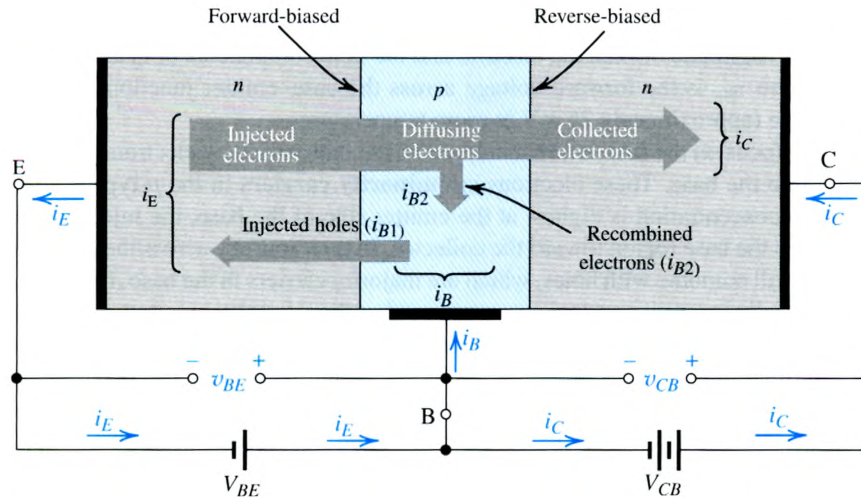


Figure 6.3 Simplified current flow in an *npn* transistor biased to operate in the active mode.

conducted by charge carriers of one type only: electrons in *n*-channel devices or holes in *p*-channel devices. In earlier days, some people referred to FETs as unipolar devices.)

6.1.2 Operation of the *npn* Transistor in the Active Mode

Of the three modes of operation of the BJT, the active mode is the most important. Therefore, we begin our study of the BJT by considering its physical operation in the active mode. (Refer to Section 3.5 if you need a refresher on the operation of the *pn* junction under forward-bias conditions: that material will come in handy here.) Figure 6.3 shows an *npn* transistor biased to operate in the active mode. Two external voltage sources are used to establish the required bias conditions for active-mode operation. The voltage V_{BE} causes the *p*-type base to be higher in potential than the *n*-type emitter, thus forward biasing the emitter–base junction. The collector–base voltage V_{CB} causes the *n*-type collector to be at a higher potential than the *p*-type base, thus reverse biasing the collector–base junction.

Current Flow The forward bias on the emitter–base junction will cause current to flow across it. This current will consist of two components: electrons injected from the emitter into the base, and holes injected from the base into the emitter. As you will see shortly, we want the first component (current caused by electrons flowing from emitter to base) to be much larger than the second component (current caused by holes flowing from base to emitter). We can accomplish this by fabricating the device with a heavily doped emitter and a lightly doped base; that is, the device is designed to have a high density of electrons in the emitter and a low density of holes in the base.

The current flowing across the emitter–base junction constitutes the emitter current i_E, as shown in Fig. 6.3. The direction of i_E is "out of" the emitter lead, which, following the usual conventions, is in the direction of the positive-charge flow (hole current) and opposite to the direction of the negative-charge flow (electron current), with the emitter current i_E being equal to the sum of these two components. However, since the electron component is much larger than the hole component, the emitter current is dominated by the electron component.

From our study in Section 3.5 of the current flow across a forward-biased *pn* junction, we know that the magnitude of both the electron and the hole components of i_E are proportional to e^{v_{BE}/V_T}, where v_{BE} is the forward voltage across the base–emitter junction and V_T is the thermal voltage (approximately 25 mV at room temperature).

Let's now focus on the first current component, the one carried by electrons injected from the emitter into the base. These electrons are **minority carriers** in the *p*-type base region. Because their concentration is highest at the emitter side of the base, the injected electrons *diffuse* through the base region toward the collector. In their journey across the base, some of the electrons will combine with holes, which are majority carriers in the base. However, since the base is usually very thin and, as we mentioned earlier, lightly doped, the proportion of electrons that are "lost" through this **recombination process** is quite small. For this reason, most of the diffusing electrons reach the boundary of the collector–base depletion region. Because the collector is more positive than the base (by the reverse-bias voltage v_{CB}), these successful electrons are swept across the CBJ depletion region into the collector. They will thus get collected and constitute the collector current i_C.

The Collector Current From what we have just seen, the collector current is carried by the electrons that reach the collector region. Its direction is opposite to that of the flow of electrons, and thus into the collector terminal. Its magnitude is proportional to e^{v_{BE}/V_T}:

$$i_C = I_S e^{v_{BE}/V_T} \tag{6.1}$$

where the constant of proportionality I_S, as in the case of the diode, is called the **saturation current** and is a transistor parameter. We will have more to say about I_S shortly.

Here it is important to note that i_C is independent of the value of v_{CB}. As long as the collector is positive with respect to the base, the electrons that reach the collector side of the base region are swept into the collector and will register as collector current.

The Base Current Figure 6.3 shows that the base current i_B is composed of two components. The first component i_{B1} is due to the holes injected from the base region into the emitter region. This current component is proportional to e^{v_{BE}/V_T}. The second component of base current, i_{B2}, is due to holes that have to be supplied by the external circuit in order to replace the holes lost from the base through the recombination process. Because i_{B2} is proportional to the number of electrons injected into the base, it is also proportional to e^{v_{BE}/V_T}. Thus the total base current, $i_B = i_{B1} + i_{B2}$, is proportional to e^{v_{BE}/V_T}, and can be expressed as a fraction of the collector current i_C:

$$i_B = \frac{i_C}{\beta} \tag{6.2}$$

That is,

$$i_B = \left(\frac{I_S}{\beta}\right) e^{v_{BE}/V_T} \tag{6.3}$$

where β is a transistor parameter.

For modern *npn* transistors, β is in the range 50 to 200, but it can be as high as 1000 for special devices. For reasons that will become clear later, the parameter β is called the **common-emitter current gain**.

The above description indicates that the value of β is highly influenced by two factors: the width of the base region, W, and the relative dopings of the base region and the emitter

region, N_A/N_D. To obtain a high β (which is what we want since β represents a gain parameter) the base should be thin (W small) and lightly doped and the emitter heavily doped (making N_A/N_D small). For modern integrated circuit fabrication technologies, W is in the nanometer range.

The Emitter Current Since the current that enters a transistor must leave it, it can be seen from Fig. 6.3 that the emitter current i_E is equal to the sum of the collector current i_C and the base current i_B:

$$i_E = i_C + i_B \tag{6.4}$$

Using Eqs. (6.2) and (6.4) gives

$$i_E = \frac{\beta + 1}{\beta} i_C \tag{6.5}$$

That is,

$$i_E = \frac{\beta + 1}{\beta} I_S e^{v_{BE}/V_T} \tag{6.6}$$

Alternatively, we can express Eq. (6.5) in the form

$$i_C = \alpha i_E \tag{6.7}$$

where the constant α is related to β by

$$\alpha = \frac{\beta}{\beta + 1} \tag{6.8}$$

Thus the emitter current in Eq. (6.6) can be written

$$i_E = (I_S/\alpha) e^{v_{BE}/V_T} \tag{6.9}$$

Finally, we can use Eq. (6.8) to express β in terms of α, that is,

$$\beta = \frac{\alpha}{1 - \alpha} \tag{6.10}$$

We can see from Eq. (6.8) that α is a constant (for a particular transistor) that is less than but very close to one. For instance, if $\beta = 100$, then $\alpha \simeq 0.99$. Equation (6.10) reveals an important fact: Small changes in α correspond to very large changes in β. This mathematical observation manifests itself physically, with the result that transistors of the same type may have widely different values of β. For reasons that will become clear later, α is called the **common-base current gain**.

Minority-Carrier Distribution We can better understand the physical operation of the BJT by considering the distribution of minority charge carriers in the base and the emitter. Figure 6.4 shows the profiles of the concentration of electrons in the base and holes in the emitter of an *npn* transistor operating in the active mode. Notice that since the doping concentration in the emitter, N_D, is much higher than the doping concentration in the base,

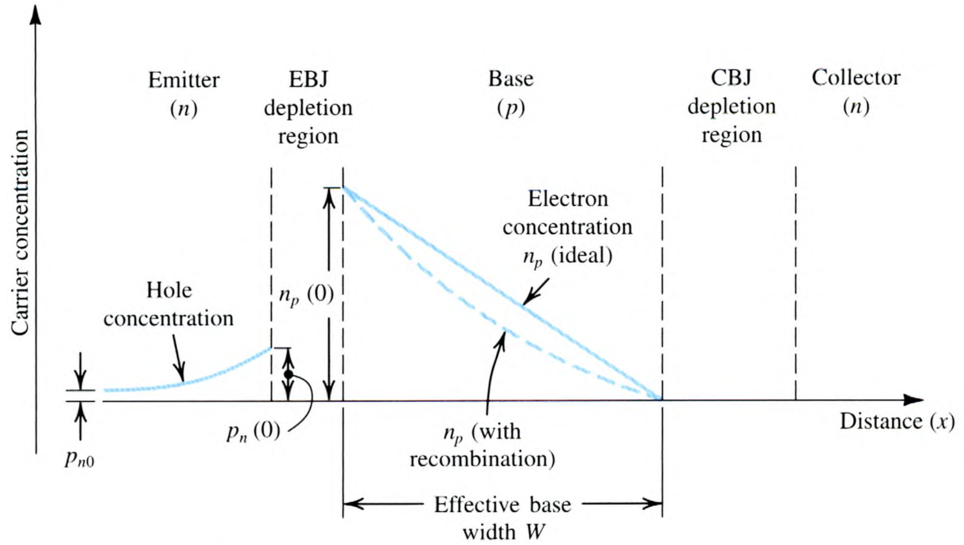


Figure 6.4 Profiles of minority-carrier concentrations in the base and in the emitter of an *npn* transistor operating in the active mode: $v_{BE} > 0$ and $v_{CB} \geq 0$.

N_A, the concentration of electrons injected from emitter to base, $n_p(0)$, is much higher than the concentration of holes injected from the base to the emitter, $p_n(0)$. Both quantities are proportional to e^{v_{BE}/V_T},

$$n_p(0) = n_{p0}\, e^{v_{BE}/V_T} \tag{6.11}$$

where n_{p0} is the thermal-equilibrium value of the minority-carrier (electron) concentration in the base region.

Next, notice that because the base is very thin, the concentration of excess electrons decays almost linearly (as opposed to the usual exponential decay we see with the excess holes in the emitter region). Furthermore, the reverse bias on the collector–base junction causes the concentration of excess electrons at the collector side of the base to be zero. (Remember that electrons that reach that point are swept into the collector.)

The tapered profile of the minority-carrier concentration (Fig. 6.4) causes the electrons injected into the base to diffuse through the base region toward the collector. This electron diffusion current I_n is directly proportional to the slope of the straight-line concentration profile,

$$\begin{aligned} I_n &= A_E q D_n \frac{dn_p(x)}{dx} \\ &= A_E q D_n \left(-\frac{n_p(0)}{W}\right) \end{aligned} \tag{6.12}$$

where A_E is the cross-sectional area of the base–emitter junction (in the direction perpendicular to the page), q is the magnitude of the electron charge, D_n is the electron diffusivity in the base, and W is the effective width of the base. Notice that the negative slope of the minority-carrier concentration results in a negative current I_n across the base; in other words, I_n flows from right to left (in the negative direction of x), which corresponds to the usual convention, opposite to the direction of electron flow.

The recombination in the base region, though slight, causes the excess minority-carrier concentration profile to deviate from a straight line and take the slightly concave shape indicated by the broken line in Fig. 6.4. The slope of the concentration profile at the EBJ is slightly higher than that at the CBJ, with the difference accounting for the small number of electrons lost in the base region through recombination.

Finally, we have the collector current $i_C = I_n$, which gives a negative value for i_C, indicating that i_C flows in the negative direction of the x axis (i.e., from right to left). Since we take this to be the positive direction of i_C, we can drop the negative sign in Eq. (6.12). Doing this and substituting for $n_p(0)$ from Eq. (6.11), we can express the collector current i_C as

$$i_C = I_S e^{v_{BE}/V_T}$$

where the saturation current I_S is given by

$$I_S = A_E q D_n n_{p0}/W$$

Substituting $n_{p0} = n_i^2/N_A$, where n_i is the intrinsic carrier density and N_A is the doping concentration in the base, we can express I_S as

$$I_S = \frac{A_E q D_n n_i^2}{N_A W} \tag{6.13}$$

The saturation current I_S is inversely proportional to the base width W and is directly proportional to the area of the EBJ. Typically I_S is in the range of 10^{-12} A to 10^{-18} A (depending on the size of the device). Because I_S is proportional to n_i^2, it is a strong function of temperature, approximately doubling for every 5°C rise in temperature. (For the dependence of n_i^2 on temperature, refer to Eq. 3.2.)

Since it is directly proportional to the junction area (i.e., the device size), I_S is also called the **scale current**. Two transistors that are identical except that one has an EBJ area, say, twice that of the other will have saturation currents with that same ratio (i.e., 2). So for the same value of v_{BE} the larger device will have a collector current twice that of the smaller device. This concept is often used in integrated-circuit design.

Recapitulation and Equivalent-Circuit Models We have presented a first-order model for the operation of the *npn* transistor in the active mode. Basically, the forward-bias voltage v_{BE} causes an exponentially related current i_C to flow in the collector terminal. The collector current i_C is independent of the collector voltage as long as the collector–base junction remains reverse biased; i.e., $v_{CB} \geq 0$. Thus in the active mode the collector terminal behaves as an ideal constant-current source where the value of the current is determined by v_{BE}. The base current i_B is a factor $1/\beta$ of the collector current, and the emitter current is equal to the sum of the collector and base currents. Since i_B is much smaller than i_C (i.e., $\beta \gg 1$), i_E is roughly equal to i_C. More precisely, the collector current is a fraction α of the emitter current, with α smaller than, but close to, unity.

We can represent this first-order model with the equivalent circuit shown in Fig. 6.5(a). Here, diode D_E has a scale current I_{SE} equal to (I_S/α), so it provides a current i_E related to v_{BE} according to Eq. (6.9). The current of the controlled source, which is equal to the collector current, is controlled by v_{BE} according to the exponential relationship indicated, a restatement of Eq. (6.1). This model is in essence a nonlinear voltage-controlled current source. We can convert it to the current-controlled current-source model shown in Fig. 6.5(b) by expressing the current of the controlled source as αi_E. Note that this model is also nonlinear because

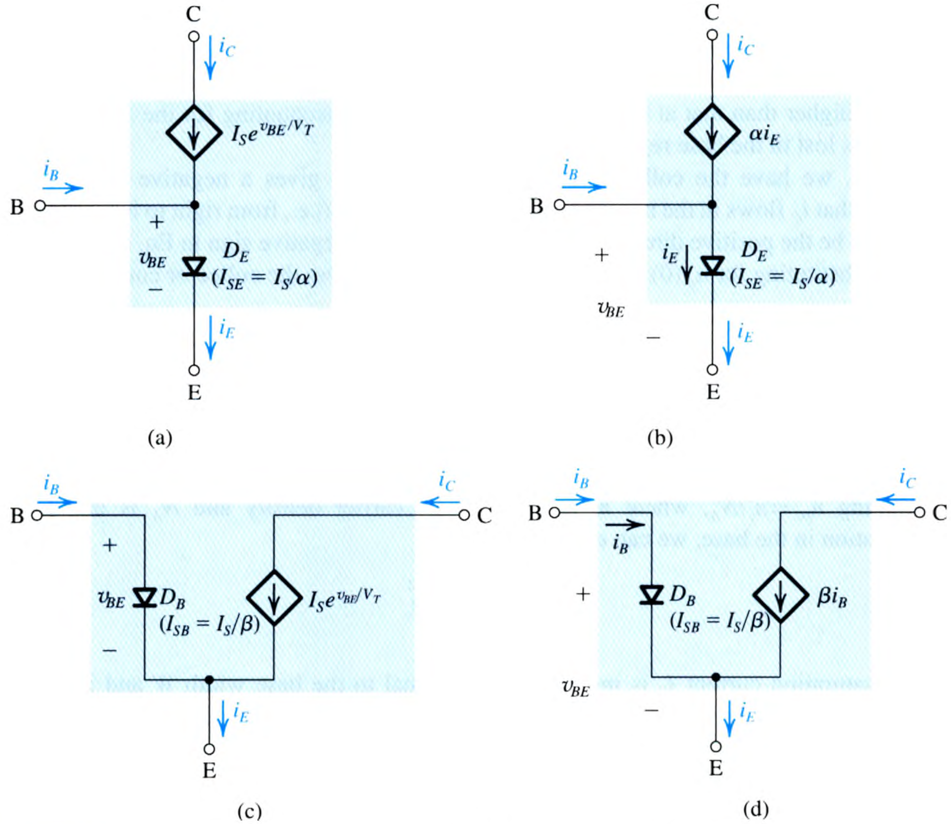


Figure 6.5 Large-signal equivalent-circuit models of the *npn* BJT operating in the active mode.

of the exponential relationship of the current i_E through diode D_E and the voltage v_{BE}. This model shows us that if the transistor is used as a two-port network with the input port between E and B and the output port between C and B (i.e., with B as a common terminal), then the current gain observed (i_C/i_E) is equal to α. This is why α is called the common-base current gain.

We can use two other equivalent-circuit models, shown in Figs. 6.5(c) and (d), to represent the operation of the BJT. The model of Fig. 6.5(c) is essentially a voltage-controlled current source. However, here diode D_B conducts the base current, so its current scale factor is I_S/β, resulting in the i_B–v_{BE} relationship given in Eq. (6.3). By simply expressing the collector current as βi_B we obtain the current-controlled current-source model shown in Fig. 6.5(d). From this latter model we observe that if the transistor is used as a two-port network with the input port between B and E and the output port between C and E (i.e., with E as the common terminal), then the current gain observed (i_C/i_B) is equal to β. This is why β is called the common-emitter current gain.

Finally, we note that the models in Fig. 6.5 apply for any positive value of v_{BE}. That is, unlike the models we will discuss in Chapter 7, here there is no limitation on the size of v_{BE}, which is why these models are referred to as **large-signal models**.

Example 6.1

An *npn* transistor having $I_S = 10^{-15}$ A and $\beta = 100$ is connected as follows: The emitter is grounded, the base is fed with a constant-current source supplying a dc current of 10 μA, and the collector is connected to a 5-V dc supply via a resistance R_C of 3 kΩ. Assuming that the transistor is operating in the active mode, find V_{BE} and V_{CE}. Use these values to verify active-mode operation. Replace the current source with a resistance connected from the base to the 5-V dc supply. What resistance value is needed to result in the same operating conditions?

Solution

If the transistor is operating in the active mode, it can be represented by any of the four equivalent-circuit models shown in Fig. 6.5. Because the emitter is grounded, the models in Figs. 6.5(c) and 6.5(d) would both be suitable. Since we know the base current I_B, the model in Fig. 6.5(d) is the most suitable.

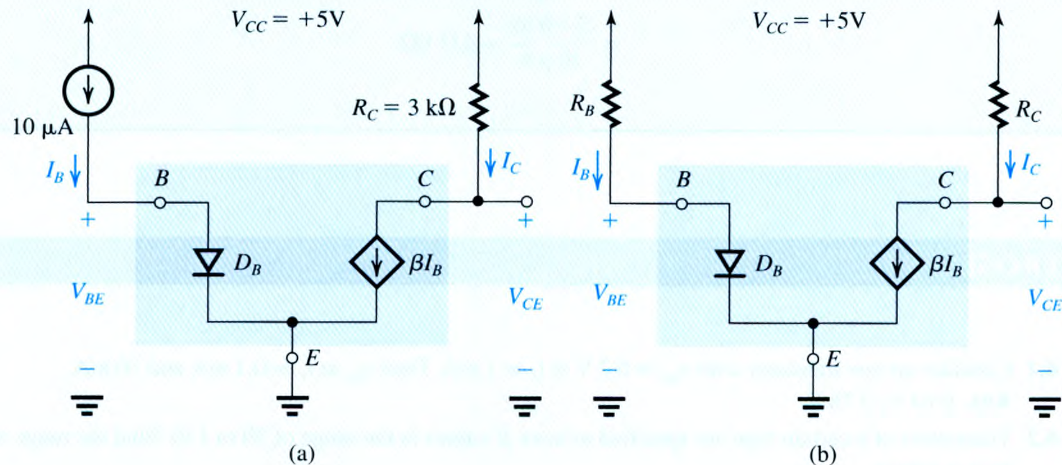


Figure 6.6 Circuits for Example 6.1.

Figure 6.6(a) shows the circuit as described with the transistor represented by the model of Fig. 6.5(d). We can determine V_{BE} from the exponential characteristic of D_B as follows:

$$V_{BE} = V_T \ln \frac{I_B}{I_S/\beta}$$

$$= 25 \ln \left(\frac{10 \times 10^{-6}}{10^{-17}} \right)$$

$$= 690 \text{ mV} = 0.69 \text{ V}$$

Example 6.1 *continued*

Next we find the value of V_{CE} from

$$V_{CE} = V_{CC} - R_C I_C$$

where

$$I_C = \beta I_B = 100 \times 10 \times 10^{-6} = 10^{-3}\ \text{A} = 1\ \text{mA}$$

Thus,

$$V_{CE} = 5 - 3 \times 1 = +2\ \text{V}$$

Since V_C (at +2 V) is higher than V_B (at 0.69 V), the transistor is indeed operating in the active mode.

Now, replacing the 10-μA current source with a resistance R_B connected from the base to the 5-V dc supply V_{CC}, as in Fig. 6.6(b), the value of R_B must be

$$R_B = \frac{V_{CC} - V_{BE}}{I_B}$$

$$= \frac{5 - 0.69}{10\ \mu\text{A}} = 431\ \text{k}\Omega$$

EXERCISES

6.1 Consider an *npn* transistor with $v_{BE} = 0.7$ V at $i_C = 1$ mA. Find v_{BE} at $i_C = 0.1$ mA and 10 mA.
Ans. 0.64 V; 0.76 V

6.2 Transistors of a certain type are specified to have β values in the range of 50 to 150. Find the range of their α values.
Ans. 0.980 to 0.993

6.3 An *npn* BJT in a particular circuit has a base current of 14.46 μA, an emitter current of 1.460 mA, and a base–emitter voltage of 0.7 V. For these conditions, calculate α, β, and I_S.
Ans. 0.99; 100; 10^{-15} A

6.4 Calculate β for two transistors for which $\alpha = 0.99$ and 0.98. For collector currents of 10 mA, find the base current of each transistor.
Ans. 99; 49; 0.1 mA; 0.2 mA

6.5 A transistor for which $I_S = 10^{-16}$ A and $\beta = 100$ is conducting a collector current of 1 mA. Find v_{BE}. Also, find I_{SE} and I_{SB} for this transistor.
Ans. 747.5 mV; 1.01×10^{-16} A; 10^{-18} A

6.6 For the circuit in Fig. 6.6(a) analyzed in Example 6.1, find the maximum value of R_C that will still result in active-mode operation.
Ans. 4.31 kΩ

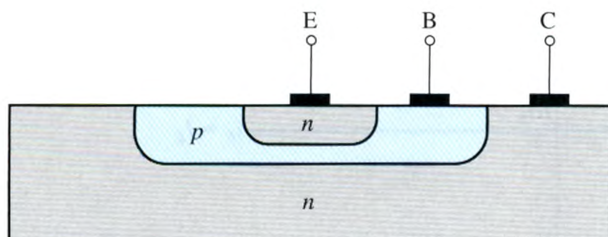


Figure 6.7 Cross section of an *npn* BJT.

6.1.3 Structure of Actual Transistors

Figure 6.7 shows a more realistic (but still simplified) cross section of an *npn* BJT. Notice that the collector virtually surrounds the emitter region, making it difficult for the electrons injected into the thin base to escape being collected. In this way, the resulting α is close to unity and β is large. Notice also that the device is *not* symmetrical, so the emitter and collector cannot be interchanged.[1] For more detail on the physical structure of actual devices, see Appendix A.

The structure in Fig. 6.7 also shows that the CBJ has a much larger area than the EBJ. Thus the CB diode D_C has a saturation current I_{SC} that is much larger than the saturation current of the EB diode D_E. Typically, I_{SC} is 10 to 100 times larger than I_{SE} (remember $I_{SE} = I_S/\alpha \simeq I_S$).

EXERCISE

6.7 A particular transistor has $I_S = 10^{-15}$ A and $\alpha \simeq 1$. If the CBJ area is 100 times the area of the EBJ, find the collector scale current I_{SC}.
Ans. 10^{-13} A

6.1.4 Operation in the Saturation Mode[2]

As we mentioned earlier, for the BJT to operate in the active mode, the CBJ must be reverse biased. So far, we have stated this condition for the *npn* transistor as $v_{CB} \geq 0$. However, we know that a *pn* junction does not effectively become forward biased until the forward voltage

[1]If the emitter and collector are reversed—that is, the CBJ is forward biased and the EBJ is reverse biased—the device operates in a mode called the "reverse-active mode." The resulting values of α and β, denoted α_R and β_R (with R denoting reverse), are much lower than the values of α and β, respectively, obtained in the "forward"-active mode discussed above. Hence, the reverse-active mode has no practical application. The MOSFET, on the other hand, being a perfectly symmetrical device, can operate equally well with its drain and source terminals interchanged.

[2]Saturation means something completely different in a BJT and in a MOSFET. The saturation mode of operation of the BJT is analogous to the triode region of operation of the MOSFET, while the saturation region of operation of the MOSFET corresponds to the active mode of BJT operation.

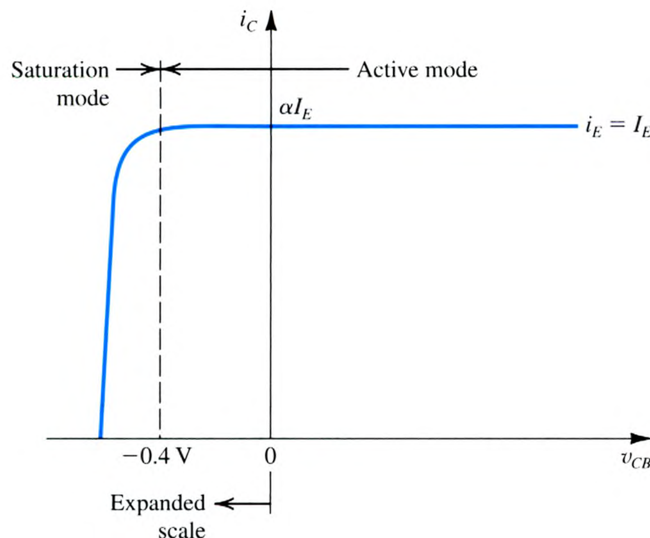


Figure 6.8 The i_C–v_{CB} characteristic of an *npn* transistor fed with a constant emitter current I_E. The transistor enters the saturation mode for $v_{CB} < -0.4$ V, and the collector current diminishes.

across it exceeds approximately 0.4 V. It follows that we can maintain active-mode operation of an *npn* transistor for negative v_{CB} down to approximately −0.4 V. This is illustrated in Fig. 6.8, which is a sketch of i_C versus v_{CB} for an *npn* transistor operated with a constant emitter current I_E. As we would expect, i_C is independent of v_{CB} in the active mode, a situation that extends for v_{CB} going negative to approximately −0.4 V. Below this value of v_{CB}, the CBJ begins to conduct sufficiently that the transistor leaves the active mode and enters the saturation mode, where i_C decreases.

To see why i_C decreases in saturation, we can build a model for the saturated *npn* transistor as follows. We augment the model of Fig. 6.5(c) with the forward-conducting CBJ diode D_C, as shown in Fig. 6.9. Notice that the current i_{BC} will subtract from the controlled-source current, resulting in the reduced collector current i_C given by

$$i_C = I_S e^{v_{BE}/V_T} - I_{SC} e^{v_{BC}/V_T} \tag{6.14}$$

where I_{SC} is the saturation current for D_C and is related to I_S by the ratio of the areas of the CBJ and the EBJ. The second term in Eq. (6.14) will play an increasing role as v_{BC} exceeds 0.4 V or so, causing i_C to decrease and eventually reach zero.

Figure 6.9 also indicates that in saturation the base current will increase to the value

$$i_B = (I_S/\beta) e^{v_{BE}/V_T} + I_{SC} e^{v_{BC}/V_T} \tag{6.15}$$

Equations (6.14) and (6.15) can be combined to obtain the ratio i_C/i_B for a saturated transistor. Note that this ratio is *lower* than the value of β. Furthermore, the ratio decreases as v_{BC} is increased and the transistor is driven deeper into saturation. Because we can set i_C/i_B of a saturated transistor to any desired value lower than β by adjusting v_{BC}, this ratio is known as **forced** β and denoted β_{forced},

$$\beta_{\text{forced}} = \left.\frac{i_C}{i_B}\right|_{\text{saturation}} \leq \beta \tag{6.16}$$

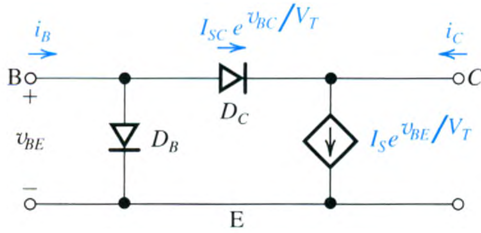


Figure 6.9 Modeling the operation of an *npn* transistor in saturation by augmenting the model of Fig. 6.5(c) with a forward-conducting diode D_C. Note that the current through D_C increases i_B and reduces i_C.

As we will show, in analyzing a circuit we can determine whether the BJT is in the saturation mode by either of the following two tests:

1. Is the CBJ forward biased by more than 0.4 V?
2. Is the ratio i_C/i_B lower than β?

The collector-to-emitter voltage v_{CE} of a saturated transistor can be found from Fig. 6.9 as the difference between the forward-bias voltages of the EBJ and the CBJ,

$$V_{CE\text{sat}} = V_{BE} - V_{BC} \tag{6.17}$$

Remember that the CBJ has a much larger area than the EBJ. As a result, V_{BC} will be smaller than V_{BE} by 0.1 to 0.3 V. Thus,

$$V_{CE\text{sat}} \simeq 0.1 \text{ to } 0.3 \text{ V}$$

Typically we will assume that a transistor at the *edge of saturation* has $V_{CE\text{sat}} = 0.3$ V, while a transistor *deep in saturation* has $V_{CE\text{sat}} = 0.2$ V.

EXERCISES

6.8 Use Eq. (6.14) to show that i_C reaches zero at

$$V_{CE} = V_T \ln(I_{SC}/I_S)$$

Calculate V_{CE} for a transistor whose CBJ has 100 times the area of the EBJ.
Ans. 115 mV

6.9 Use Eqs. (6.14), (6.15), and (6.16) to show that a BJT operating in saturation with $V_{CE} = V_{CE\text{sat}}$ has a forced β given by

$$\beta_{\text{forced}} = \beta \frac{e^{V_{CE\text{sat}}/V_T} - I_{SC}/I_S}{e^{V_{CE\text{sat}}/V_T} + \beta I_{SC}/I_S}$$

Find β_{forced} for $\beta = 100$, $I_{SC}/I_S = 100$, and $V_{CE\text{sat}} = 0.2$ V.
Ans. 22.2

6.1.5 The *pnp* Transistor

The *pnp* transistor operates in a manner similar to that of the *npn* device. Figure 6.10 shows a *pnp* transistor biased to operate in the active mode. Here the voltage V_{EB} causes the *p*-type

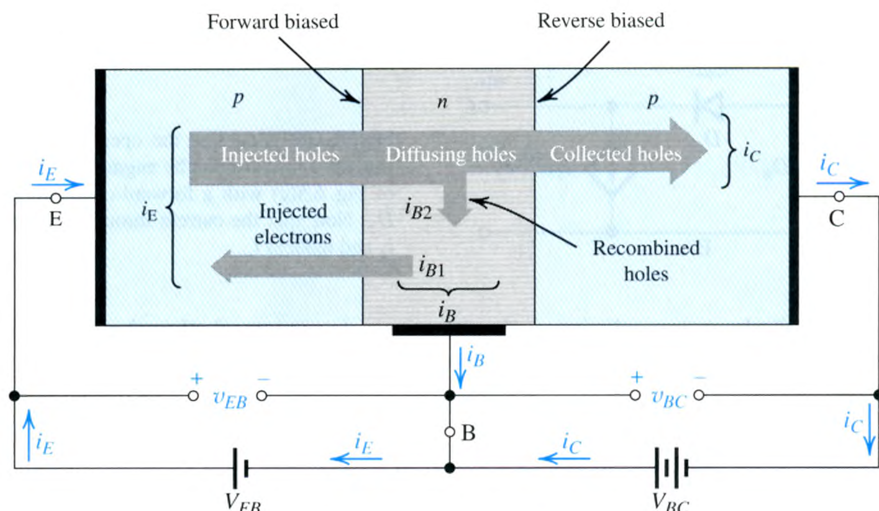


Figure 6.10 Current flow in a *pnp* transistor biased to operate in the active mode.

emitter to have higher potential than the *n*-type base, thus forward biasing the emitter–base junction. The collector–base junction is reverse biased by the voltage V_{BC}, which keeps the *p*-type collector lower in potential than the *n*-type base.

Unlike the *npn* transistor, current in the *pnp* device is mainly conducted by holes injected from the emitter into the base as a result of the forward-bias voltage v_{EB}. Since the component of emitter current contributed by electrons injected from base to emitter is kept small by using a lightly doped base, most of the emitter current is due to holes. The electrons injected from base to emitter give rise to the first component of base current, i_{B1}. Also, a number of the holes injected into the base recombine with the majority carriers in the base (electrons) and will thus be lost. The disappearing base electrons have to be replaced from the external circuit, giving rise to the second component of base current, i_{B2}. The holes that succeed in reaching the boundary of the depletion region of the collector–base junction are attracted by the negative voltage on the collector. Thus these holes are swept across the depletion region into the collector and appear as collector current.

It is clear from the this description that the current–voltage relationship of the *pnp* transistor will be identical to that of the *npn* transistor except that v_{BE} has to be replaced by v_{EB}. Also, the large-signal, active-mode operation of the *pnp* transistor can be modeled by any of four equivalent circuits similar to those for the *npn* transistor in Fig. 6.5. Two of these four circuits are shown in Fig. 6.11. Finally, the *pnp* transistor can operate in the saturation mode in a manner analogous to that described for the *npn* device.

EXERCISES

6.10 Consider the model in Fig. 6.11(a) applied in the case of a *pnp* transistor whose base is grounded, whose emitter is fed by a constant-current source of 2 mA, and whose collector is connected to a −10-V dc supply. Find the emitter voltage, the base current, and the collector current if $\beta = 50$ and $I_S = 10^{-14}$ A.

Ans. 0.650 V; 39.2 μA; 1.96 mA

6.11 For a *pnp* transistor having $I_S = 10^{-11}$ A and $\beta = 100$, calculate v_{EB} for $i_C = 1.5$ A.
Ans. 0.643 V

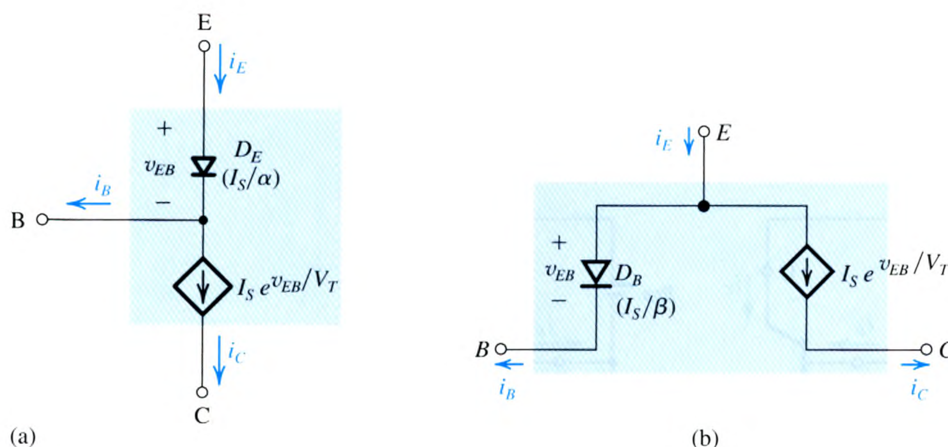


Figure 6.11 Two large-signal models for the *pnp* transistor operating in the active mode.

THE INVENTION OF THE BJT

The first working transistor was demonstrated at Bell Labs in late 1947 by John Bardeen and Walter Brattain, who were part of a team led by William Shockley. Made of germanium, the device became known as a point-contact transistor and operated on the field-effect principle. Within a few weeks, however, Shockley wrote a complete description of the bipolar junction transistor (BJT) and filed for a U.S. patent with the title "Circuit Element Utilizing Semiconductor Material."

BJTs dominated the electronics world from the early 1950s to the mid-1970s, when MOSFETs took over the leading position. In 1956, Shockley, Bardeen, and Brattain shared the Nobel Prize in Physics for the discovery of the transistor effect.

6.2 Current–Voltage Characteristics

6.2.1 Circuit Symbols and Conventions

The physical structure we have used so far to explain transistor operation is awkward to use when drawing the schematic of a multitransistor circuit. Fortunately, there is a very descriptive and convenient circuit symbol for the BJT. Figures 6.12(a) and (b) show the symbol for the *npn* and the *pnp* transistors, respectively. In both symbols the emitter is distinguished by an arrowhead. This distinction is important because, as we saw in the last section, BJTs are not symmetric devices.

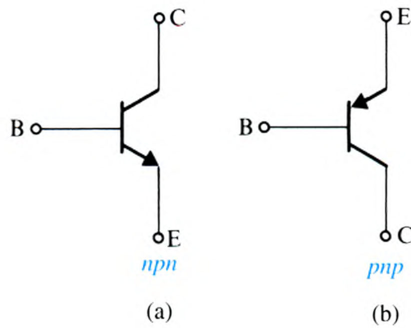


Figure 6.12 Circuit symbols for BJTs.

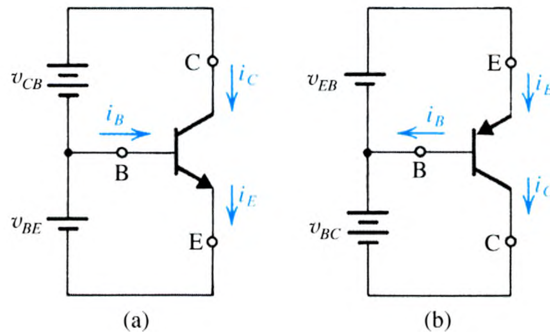


Figure 6.13 Voltage polarities and current flow in transistors operating in the active mode.

The polarity of the device—*npn* or *pnp*—is indicated by the direction of the arrowhead. It points in the direction of normal current flow in the emitter, which is also the forward direction of the base–emitter junction. Since we have adopted a drawing convention by which currents flow from top to bottom, we will always draw *pnp* transistors with their emitters on top, as shown in Fig. 6.12(b).

Figure 6.13 shows *npn* and *pnp* transistors connected to dc sources so that they operate in the active mode. Figure 6.13 also shows the reference and actual directions of current flow throughout the transistor. Our convention is to take the reference direction to coincide with the normal direction of current flow. Normally, then, we should not encounter a negative value for i_E, i_B, or i_C.

The convenience of this circuit-drawing convention should be obvious from Fig. 6.13. Currents flow from top to bottom and voltages are higher at the top and lower at the bottom. The arrowhead on the emitter also indicates the polarity of the emitter–base voltage that should be applied in order to forward-bias the emitter–base junction. Just a glance at the circuit symbol of the *pnp* transistor, for example, tells us that we should make the emitter higher in voltage than the base (by v_{EB}) in order to make current flow into the emitter (downward). Note that the symbol v_{EB} means the voltage by which the emitter (E) is higher than the base (B). Thus for a *pnp* transistor operating in the active mode, v_{EB} is positive; in an *npn* transistor, v_{BE} is positive.

From the discussion in Section 6.1 it follows that an *npn* transistor whose EBJ is forward biased (usually, $v_{BE} \simeq 0.7$ V) will operate in the active mode *as long as the collector voltage*

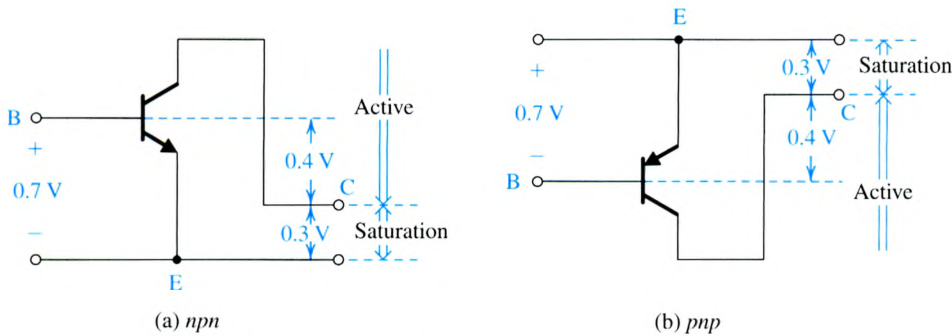


Figure 6.14 Graphical representation of the conditions for operating the BJT in the active mode and in the saturation mode.

Table 6.2 Summary of the BJT Current–Voltage Relationships in the Active Mode

$$i_C = I_S e^{v_{BE}/V_T}$$

$$i_B = \frac{i_C}{\beta} = \left(\frac{I_S}{\beta}\right) e^{v_{BE}/V_T}$$

$$i_E = \frac{i_C}{\alpha} = \left(\frac{I_S}{\alpha}\right) e^{v_{BE}/V_T}$$

Note: For the *pnp* transistor, replace v_{BE} with v_{EB}.

$i_C = \alpha i_E$	$i_B = (1-\alpha)i_E = \dfrac{i_E}{\beta+1}$
$i_C = \beta i_B$	$i_E = (\beta+1)i_B$
$\beta = \dfrac{\alpha}{1-\alpha}$	$\alpha = \dfrac{\beta}{\beta+1}$

$$V_T = \text{thermal voltage} = \frac{kT}{q} \simeq 25 \text{ mV at room temperature}$$

does not fall below that of the base by more than approximately 0.4 V. Otherwise, the transistor leaves the active mode and enters the saturation region.[3]

Similarly, the *pnp* transistor will operate in the active mode *if the EBJ is forward biased (usually,* $v_{EB} \simeq 0.7$ V*) and the collector voltage is not allowed to rise above that of the base by more than 0.4 V or so*. Otherwise, the CBJ becomes forward biased, and the *pnp* transistor enters the saturation region.

For greater emphasis, Fig. 6.14 illustrates the conditions for operating the BJT in the active mode and in the saturation mode. Also, for easy reference, Table 6.2 summarizes the BJT current–voltage relationships in the active mode of operation.

[3]It is interesting to contrast the active-mode operation of the BJT with the corresponding mode of operation of the MOSFET (saturation): The BJT needs a minimum v_{CE} of about 0.3 V, and the MOSFET needs a minimum v_{DS} equal to V_{OV}, which for modern technologies is in the range of 0.2 V to 0.3 V. Thus we see a great deal of similarity! Also note that reverse biasing the CBJ of the BJT corresponds to pinching off the channel of the MOSFET. This condition causes the collector current (drain current in the MOSFET) to be independent of the collector voltage (the drain voltage in the MOSFET).

The Collector–Base Reverse Current (I_{CBO}) In our discussion of current flow in transistors we ignored the small reverse currents carried by thermally generated minority carriers. Although such currents can be safely neglected in modern transistors, the reverse current across the collector–base junction deserves some mention. This current, denoted I_{CBO}, is the reverse current flowing from collector to base with the emitter open-circuited (hence the subscript O). This current is usually in the nanoampere range, a value that is many times higher than its theoretically predicted value because I_{CBO} contains a substantial leakage component. Its value is dependent on v_{CB} and depends strongly on temperature, approximately doubling for every 10°C rise.

Example 6.2

The transistor in the circuit of Fig. 6.15(a) has $\beta = 100$ and exhibits a v_{BE} of 0.7 V at $i_C = 1$ mA. Design the circuit so that a current of 2 mA flows through the collector and a voltage of +5 V appears at the collector.

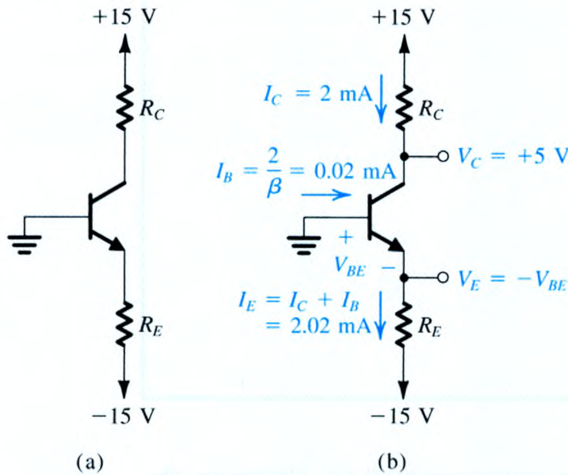


Figure 6.15 Circuit for Example 6.2.

Solution

Refer to Fig. 6.15(b). Since we are designing for $V_C = +5$ V, the CBJ will be reverse biased and the BJT will be operating in the active mode. To obtain a voltage $V_C = +5$ V, the voltage drop across R_C must be $15 - 5 = 10$ V. Now, since $I_C = 2$ mA, the value of R_C should be selected according to

$$R_C = \frac{10\text{ V}}{2\text{ mA}} = 5\text{ k}\Omega$$

Since $v_{BE} = 0.7$ V at $i_C = 1$ mA, the value of v_{BE} at $i_C = 2$ mA is

$$V_{BE} = 0.7 + V_T \ln\left(\frac{2}{1}\right) = 0.717\text{ V}$$

Since the base is at 0 V, the emitter voltage should be

$$V_E = -0.717\text{ V}$$

For $\beta = 100$, $\alpha = 100/101 = 0.99$. Thus the emitter current should be

$$I_E = \frac{I_C}{\alpha} = \frac{2}{0.99} = 2.02 \text{ mA}$$

Now the value required for R_E can be determined from

$$R_E = \frac{V_E - (-15)}{I_E}$$
$$= \frac{-0.717 + 15}{2.02} = 7.07 \text{ k}\Omega$$

This completes the design. We should note, however, that the calculations above were made with a degree of precision that is usually neither necessary nor justified in practice in view, for instance, of the expected tolerances of component values. Nevertheless, we chose to do the design precisely in order to illustrate the various steps involved.

EXERCISES

D6.12 Repeat Example 6.2 for a transistor fabricated in a modern integrated-circuit process. Such a process yields devices that exhibit larger v_{BE} at the same i_C because they have much smaller junction areas. The dc power supplies used in modern IC technologies fall in the range of 1 V to 3 V. Design a circuit similar to that shown in Fig. 6.15 except that now the power supplies are ± 1.5 V and the BJT has $\beta = 100$ and exhibits v_{BE} of 0.8 V at $i_C = 1$ mA. Design the circuit so that a current of 2 mA flows through the collector and a voltage of +0.5 V appears at the collector.

Ans. $R_C = 500\ \Omega$; $R_E = 338\ \Omega$

6.13 In the circuit shown in Fig. E6.13, the voltage at the emitter was measured and found to be –0.7 V. If $\beta = 50$, find I_E, I_B, I_C, and V_C.

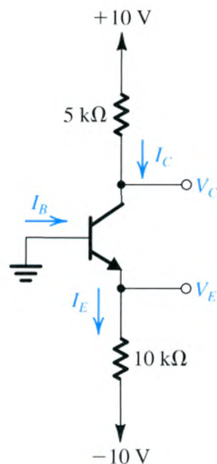


Figure E6.13

Ans. 0.93 mA; 18.2 μA; 0.91 mA; +5.45 V

6.14 In the circuit shown in Fig. E6.14, measurement indicates V_B to be +1.0 V and V_E to be +1.7 V. What are α and β for this transistor? What voltage V_C do you expect at the collector?

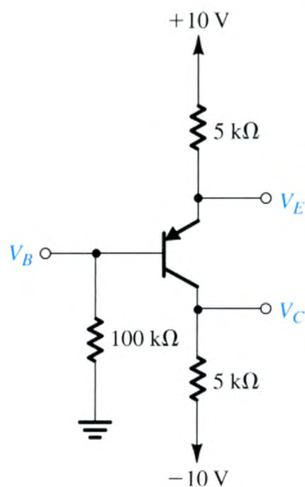


Figure E6.14

Ans. 0.994; 165; –1.75 V

Video Example VE 6.1

Design the circuit in Fig. VE6.1 to establish $I_C = 0.2$ mA and $V_C = 0.5$ V. The transistor exhibits v_{BE} of 0.8 V at $i_C = 1$ mA, and $\beta = 100$.

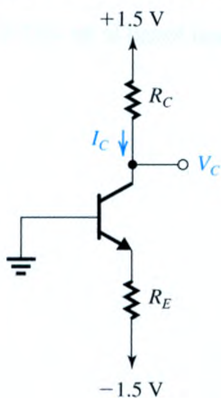


Figure VE6.1 Circuit for Video Example 6.1.



VE 6.1

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 6.34

6.2.2 Graphical Representation of Transistor Characteristics

It is sometimes useful to describe the transistor i–v characteristics graphically. Figure 6.16 shows the i_C–v_{BE} characteristic, which is the exponential relationship

$$i_C = I_S e^{v_{BE}/V_T}$$

which is identical to the diode i–v relationship. The i_E–v_{BE} and i_B–v_{BE} characteristics are also exponential but with different scale currents: I_S/α for i_E, and I_S/β for i_B. Since the constant of the exponential characteristic, $1/V_T$, is quite high ($\simeq$ 40), the curve rises very sharply. For v_{BE} smaller than about 0.5 V, the current is negligibly small.[4] Also, over most of the normal current range, v_{BE} lies in the range of 0.6 V to 0.8 V. In performing rapid first-order dc calculations, we normally assume that $V_{BE} \simeq 0.7$ V, which is similar to the approach we used in the analysis of diode circuits (Chapter 4). For a *pnp* transistor, the i_C–v_{EB} characteristic looks identical to that of Fig. 6.16, with v_{BE} replaced with v_{EB}.

As in silicon diodes, the voltage across the emitter–base junction decreases by about 2 mV for each rise of 1°C in temperature, provided the junction is operating at a constant current. Figure 6.17 illustrates this temperature dependence by depicting i_C–v_{BE} curves for an *npn* transistor at three different temperatures.

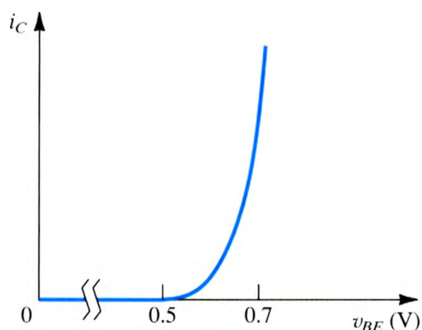


Figure 6.16 The i_C–v_{BE} characteristic for an *npn* transistor.

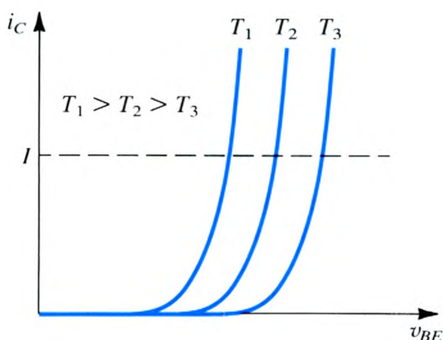


Figure 6.17 Effect of temperature on the i_C–v_{BE} characteristic. At a constant emitter current (broken line), v_{BE} changes by −2 mV/°C.

[4]The i_C–v_{BE} characteristic is the BJT's counterpart of the i_D–v_{GS} characteristic of the MOSFET. They share an important attribute: In both cases the voltage has to exceed a "threshold" for the device to conduct appreciably. In the case of the MOSFET, there is a formal threshold voltage, V_t, which lies typically in the range of 0.3 V to 0.8 V. For the BJT, there is an "apparent threshold" of approximately 0.5 V. The i_D–v_{GS} characteristic of the MOSFET is parabolic, and thus not as steep as the i_C–v_{BE} characteristic of the BJT. As will be seen in Chapter 7, this difference has a direct and significant implication for the value of transconductance g_m realized with each device.

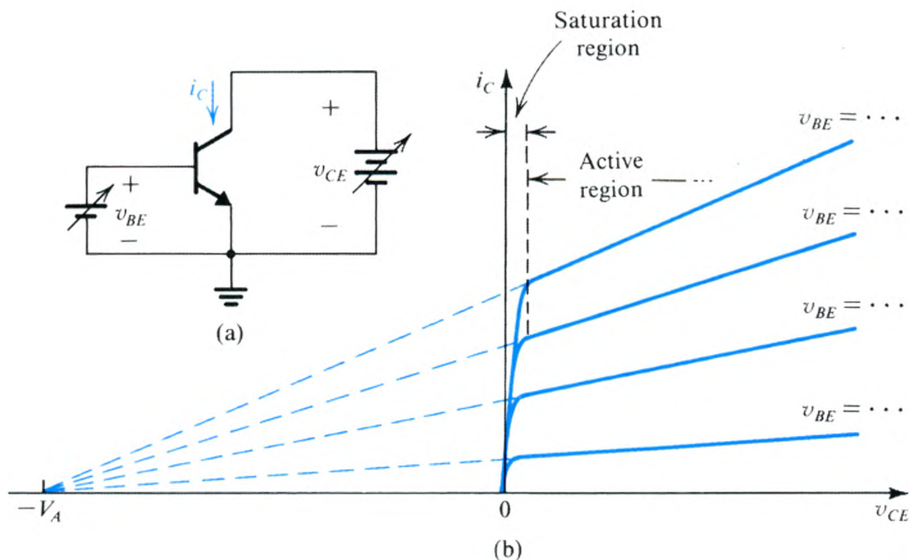


Figure 6.18 (**a**) Conceptual circuit for measuring the i_C–v_{CE} characteristics of the BJT. (**b**) The i_C–v_{CE} characteristics of a practical BJT.

EXERCISE

6.15 Consider a *pnp* transistor with $v_{EB} = 0.7$ V at $i_E = 1$ mA. Let the base be grounded, the emitter be fed by a 2-mA constant-current source, and the collector be connected to a –5-V supply through a 1-kΩ resistance. If the temperature increases by 30°C, find the changes in emitter and collector voltages. Neglect the effect of I_{CBO}.
Ans. –60 mV; 0 V

6.2.3 Dependence of i_C on the Collector Voltage—The Early Effect

In the active region, the collector current of a practical BJT shows some dependence on the collector voltage. As a result, unlike the graph shown in Fig. 6.8, the i_C–v_{CB} characteristics of a practical BJT are not perfectly horizontal straight lines. To see this dependence more clearly, consider the conceptual circuit in Fig. 6.18(a). The transistor is connected in the **common-emitter configuration**; that is, here the emitter serves as a common terminal between the input and output ports. The voltage v_{BE} can be set to any desired value by adjusting the dc source connected between base and emitter. At each value of v_{BE}, the corresponding i_C–v_{CE} characteristic curve can be measured point by point by varying the dc source connected between collector and emitter and measuring the corresponding collector current. The result is the family of i_C–v_{CE} characteristic curves shown in Fig. 6.18(b) and known as **common-emitter characteristics**.

At low values of v_{CE} (lower than about 0.3 V), as the collector voltage goes below that of the base by more than 0.4 V, the collector–base junction becomes forward biased

and the transistor leaves the active mode and enters the saturation mode. We will look at details of the i_C–v_{CE} curves in the saturation region shortly. At this time, however, we wish to examine the characteristic curves in the active region in detail. We observe that the characteristic curves, though still straight lines, have finite slope. In fact, when extrapolated, the characteristic lines meet at a point on the negative v_{CE} axis, at $v_{CE} = -V_A$. The voltage V_A, a positive number, is a parameter for the particular BJT, with typical values in the range of 10 V to 100 V. It is called the **Early voltage**, after J. M. Early, the engineering scientist who first studied this phenomenon.

At a given value of v_{BE}, increasing v_{CE} increases the reverse-bias voltage on the collector–base junction, and thus increases the width of the depletion region of this junction (refer to Fig. 6.4). This produces a decrease in the **effective base width** W. Because I_S is inversely proportional to W (Eq. 6.13), I_S will increase and i_C increases proportionally. This is the Early effect, also known as the **base-width modulation effect.**[5]

We can account for the linear dependence of i_C on v_{CE} by assuming that I_S remains constant and including the factor $(1 + v_{CE}/V_A)$ in the equation for i_C as follows:

$$i_C = I_S e^{v_{BE}/V_T}\left(1 + \frac{v_{CE}}{V_A}\right) \tag{6.18}$$

The nonzero slope of the i_C–v_{CE} straight lines indicates that the **output resistance** looking into the collector is not infinite. Rather, it is finite and defined by

$$r_o \equiv \left[\left.\frac{\partial i_C}{\partial v_{CE}}\right|_{v_{BE} = \text{constant}}\right]^{-1} \tag{6.19}$$

Using Eq. (6.18) we can show that

$$r_o = \frac{V_A + V_{CE}}{I_C} \tag{6.20}$$

where I_C and V_{CE} are the coordinates of the point at which the BJT is operating on the particular i_C–v_{CE} curve (i.e., the curve obtained for v_{BE} equal to the constant value V_{BE} at which Eq. (6.19) is evaluated). Alternatively, we can write

$$r_o = \frac{V_A}{I'_C} \tag{6.21}$$

where I'_C is the value of the collector current with the Early effect neglected; that is,

$$I'_C = I_S e^{V_{BE}/V_T} \tag{6.22}$$

It is rarely necessary to include the dependence of i_C on v_{CE} in dc bias design and analysis that is performed by hand. Such an effect, however, can be easily included in the SPICE simulation of circuit operation, which is frequently used to "fine-tune" pencil-and-paper analysis or design.

The finite output resistance r_o can have a significant effect on the gain of transistor amplifiers. This is particularly the case in integrated-circuit amplifiers, as we will see in Chapter 8. Fortunately, there are many situations in which r_o can be included relatively easily in pencil-and-paper analysis. For simplicity, however, we usually drop the prime from Eq. (6.21) and simply use $r_o = V_A/I_C$, where I_C is the collector current without the Early effect.

The output resistance r_o can be included in the circuit model of the transistor. This is illustrated in Fig. 6.19, where we show the two large-signal circuit models of a

[5] Recall that the MOSFET's counterpart is the channel-length modulation effect. These two effects are remarkably similar and have been assigned the same name, Early effect.

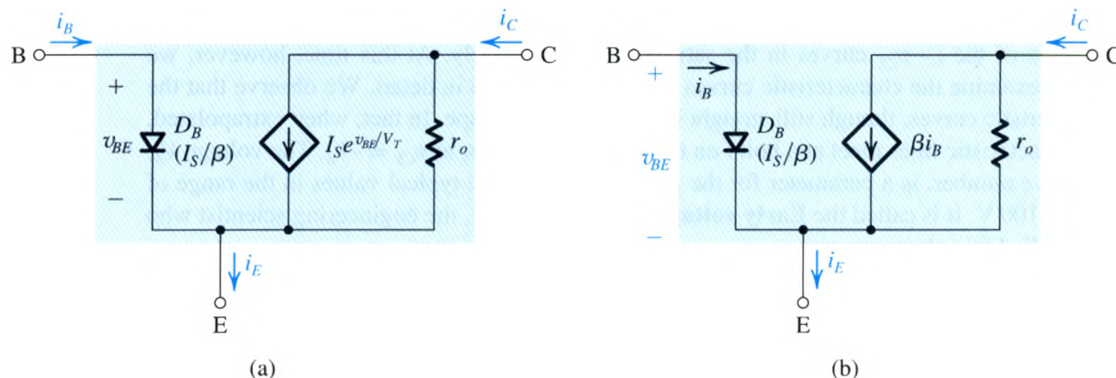


Figure 6.19 Large-signal, equivalent-circuit models of an *npn* BJT operating in the active mode in the common-emitter configuration with the output resistance r_o included.

common-emitter *npn* transistor operating in the active mode, those in Figs. 6.5(c) and (d), with the resistance r_o connected between the collector and the emitter terminals.

EXERCISES

6.16 Use the circuit model in Fig. 6.19(a) to express i_C in terms of e^{v_{BE}/V_T} and v_{CE} and thus show that this circuit is a direct representation of Eq. (6.18).

6.17 Find the output resistance of a BJT for which $V_A = 100$ V at $I_C = 0.1$, 1, and 10 mA.
Ans. 1 MΩ; 100 kΩ; 10 kΩ

6.18 Consider the circuit in Fig. 6.18(a). At $v_{CE} = 1$ V, v_{BE} is adjusted to yield a collector current of 1 mA. Then, while v_{BE} is kept constant, v_{CE} is raised to 11 V. Find the new value of I_C. For this transistor, $V_A = 100$ V.
Ans. 1.1 mA

6.2.4 An Alternative Form of the Common-Emitter Characteristics

We show another way of expressing the transistor common-emitter characteristics in Fig. 6.20. Here the base current i_B, rather than the base–emitter voltage v_{BE}, is used as a parameter. That is, each i_C–v_{CE} curve is measured with the base fed with a constant current i_B. The resulting characteristics, shown in Fig. 6.20(b), look similar to those in Fig. 6.18. Figure 6.20(c) shows an expanded view of the characteristics in the saturation region.

The Common-Emitter Current Gain β In the active region of the characteristics in Fig. 6.20(b) we have identified a particular point Q. Note that this operating point for the transistor is characterized by a base current I_B, a collector current I_C, and a collector–emitter voltage V_{CE}. The ratio I_C/I_B is the transistor β. However, there is another way to measure β:

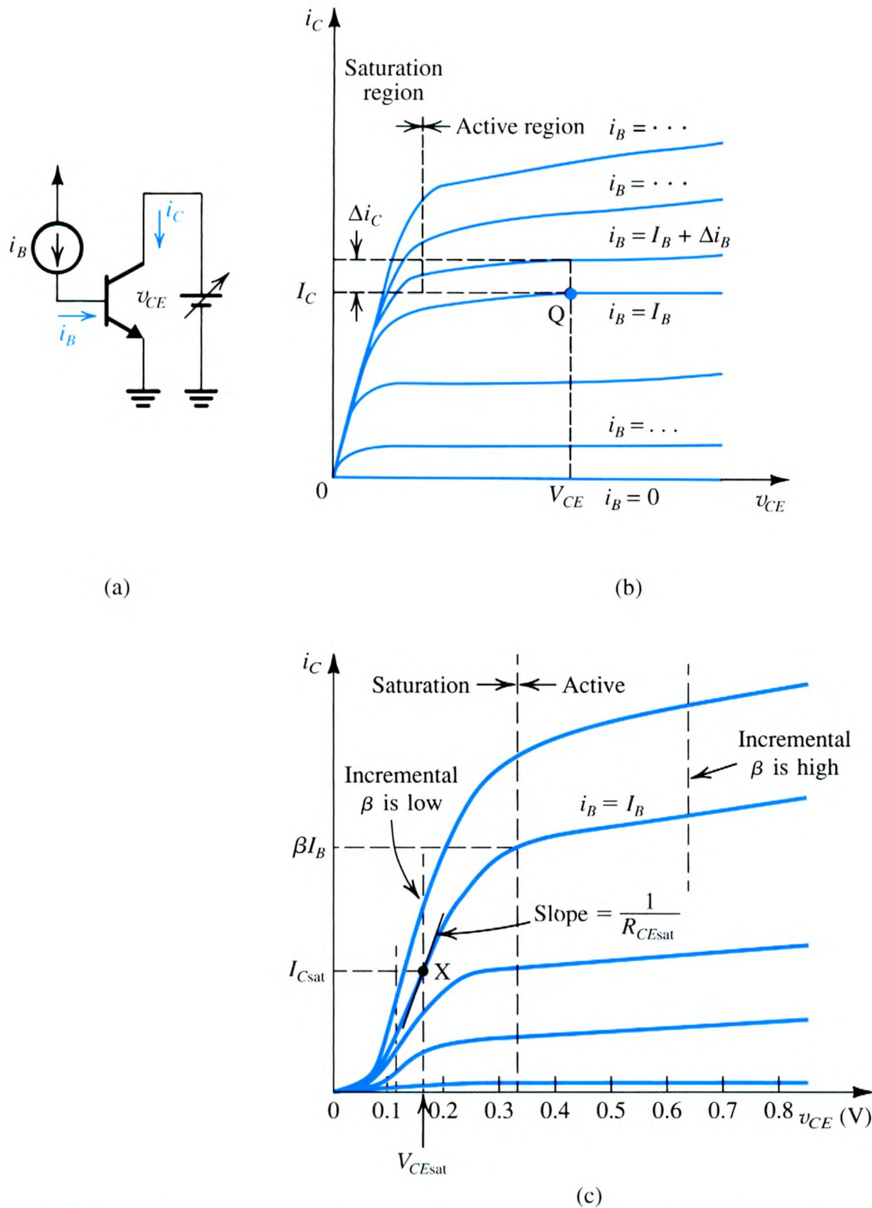


Figure 6.20 Common-emitter characteristics. (**a**) Basic CE circuit; note that in (**b**) the horizontal scale is expanded around the origin to show the saturation region in some detail. A much greater expansion of the saturation region is shown in (**c**) .

change the base current by an increment Δi_B and measure the resulting increment Δi_C, while keeping V_{CE} constant. This is illustrated in Fig. 6.20(b). The ratio $\Delta i_C/\Delta i_B$ should, according to our study so far, give an identical value for β. It turns out, however, that the latter value of β (called *incremental*, or ac, β) is a little different from the dc β (i.e., I_C/I_B). The difference,

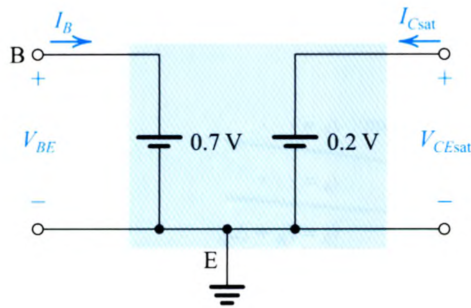


Figure 6.21 A simplified equivalent-circuit model of the saturated transistor.

however, is too subtle for our needs here. We shall use β to denote both dc and incremental values.[6]

The Saturation Voltage V_{CEsat} and Saturation Resistance R_{CEsat} Refer next to the expanded view of the common-emitter characteristics in the saturation region in Fig. 6.20(c). The "bunching together" of the curves in the saturation region implies that the incremental β is lower there than in the active region. A possible operating point in the saturation region is labeled X. It is characterized by a base current I_B, a collector current I_{Csat}, and a collector–emitter voltage V_{CEsat}. From our previous discussion of saturation, recall that $I_{Csat} = \beta_{forced} I_B$, where $\beta_{forced} < \beta$.

The i_C–v_{CE} curves in saturation are rather steep, indicating that the saturated transistor exhibits a low collector-to-emitter resistance R_{CEsat},

$$R_{CEsat} \equiv \left.\frac{\partial v_{CE}}{\partial i_C}\right|_{\substack{i_B = I_B \\ i_C = I_{Csat}}} \tag{6.23}$$

Typically, R_{CEsat} ranges from a few ohms to a few tens of ohms.

That the collector-to-emitter resistance of a saturated BJT is small should have been anticipated from the fact that between C and E we now have two forward-conducting diodes in series[7] (see also Fig. 6.9).

A simple model for the saturated BJT is shown in Fig. 6.21. Here we assume that V_{BE} is constant at approximately 0.7 V and V_{CE} is constant at approximately 0.2 V. In other words, we have neglected the small saturation resistance R_{CEsat} for the sake of making the model simple for hand calculations.

Example 6.3

For the circuit in Fig. 6.22, determine the value of the voltage V_{BB} that results in the transistor operating

(a) in the active mode with $V_{CE} = 5$ V

[6]Manufacturers of bipolar transistors use h_{FE} to denote the dc value of β and h_{fe} to denote the incremental β. These symbols come from the h-parameter description of two-port networks (see Appendix C), with the subscript $F(f)$ denoting forward and $E(e)$ denoting common emitter.

[7]In the corresponding mode of operation for the MOSFET, the triode region, the resistance between drain and source is small because it is the resistance of the continuous (non-pinched-off) channel.

(b) at the edge of saturation
(c) deep in saturation with $\beta_{\text{forced}} = 10$

For simplicity, assume that V_{BE} remains constant at 0.7 V. The transistor β is specified to be 50.

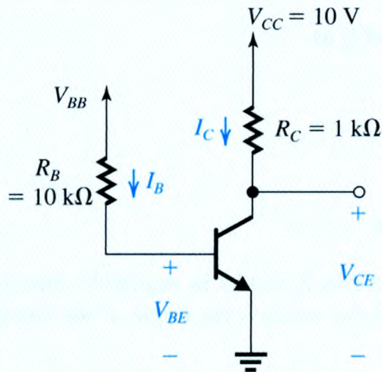


Figure 6.22 Circuit for Example 6.3.

Solution

(a) To operate in the active mode with $V_{CE} = 5$ V,

$$I_C = \frac{V_{CC} - V_{CE}}{R_C} = \frac{10 - 5}{1 \text{ k}\Omega} = 5 \text{ mA}$$

$$I_B = \frac{I_C}{\beta} = \frac{5}{50} = 0.1 \text{ mA}$$

Now we can find the required value of V_{BB}:

$$\begin{aligned} V_{BB} &= I_B R_B + V_{BE} \\ &= 0.1 \times 10 + 0.7 = 1.7 \text{ V} \end{aligned}$$

(b) Operation at the edge of saturation is obtained when $V_{CE} = 0.3$ V. Thus

$$I_C = \frac{10 - 0.3}{1} = 9.7 \text{ mA}$$

Since, at the edge of saturation, I_C and I_B are still related by β,

$$I_B = \frac{9.7}{50} = 0.194 \text{ mA}$$

The required value of V_{BB} can be determined as

$$V_{BB} = 0.194 \times 10 + 0.7 = 2.64 \text{ V}$$

(c) To operate deep in saturation,

$$V_{CE} = V_{CE\text{sat}} \simeq 0.2 \text{ V}$$

Example 6.3 *continued*

Thus,

$$I_C = \frac{10 - 0.2}{1} = 9.8 \text{ mA}$$

We then use the value of forced β to determine the required value of I_B as

$$I_B = \frac{I_C}{\beta_{\text{forced}}} = \frac{9.8}{10} = 0.98 \text{ mA}$$

and the required V_{BB} can now be found as

$$V_{BB} = 0.98 \times 10 + 0.7 = 10.5 \text{ V}$$

Notice that once the transistor is in saturation, increasing V_{BB} and thus I_B results in negligible change in I_C since V_{CEsat} will change only slightly. Thus I_C is said to *saturate*, which is the origin of the name "saturation mode of operation."

EXERCISES

6.19 Repeat Example 6.3 for $R_C = 10$ kΩ.
Ans. 0.8 V; 0.894 V; 1.68 V

6.20 For the circuit in Fig. 6.22, find V_{CE} for $V_{BB} = 0$ V.
Ans. + 10 V

6.21 For the circuit in Fig. 6.22, let V_{BB} be set to the value obtained in Example 6.3, part (a), namely, $V_{BB} = 1.7$ V. Verify that the transistor is indeed operating in the active mode. Now, while keeping V_{BB} constant, find the value to which R_C should be increased in order to obtain (a) operation at the edge of saturation and (b) operation deep in saturation with $\beta_{\text{forced}} = 10$.
Ans. (a) 1.94 kΩ; (b) 9.8 kΩ

6.3 BJT Circuits at DC

We are now ready to consider the analysis of BJT circuits to which only dc voltages are applied. In the following examples we use the simple model in which $|V_{BE}|$ of a conducting transistor is 0.7 V and $|V_{CE}|$ of a saturated transistor is 0.2 V, and we neglect the Early effect. These models are shown in Table 6.3. Better models can, of course, be used to obtain more accurate results, but analysis using these more complex models takes more

Table 6.3 Simplified Models for the Operation of the BJT in DC Circuits

	npn	*pnp*
Active EBJ: Forward Biased CBJ: Reverse Biased	I_B B, C, E + $V_{BE} \simeq 0.7$ V − βI_B + $V_{CE} \geq 0.3$ V −	E + $V_{EB} \simeq 0.7$ V − βI_B + $V_{EC} \geq 0.3$ V − B, C I_B
Saturation EBJ: Forward Biased CBJ: Forward Biased	I_B $I_C = \beta_{\text{forced}} I_B$ B, C, E + $V_{BE} \simeq 0.7$ V − + $V_{CE\text{sat}} \simeq 0.2$ V −	E + $V_{EB} \simeq 0.7$ V − + $V_{EC\text{sat}} \simeq 0.2$ V − B, C I_B $I_C = \beta_{\text{forced}} I_B$

$\beta_{\text{forced}} < \beta_{\min}$

time; more importantly, the added complexity could impede our ability to gain insight into circuit behavior. Accurate results using elaborate models can be obtained using circuit simulation with SPICE. This is almost always done in the final stages of a design and certainly before circuit fabrication. Computer simulation, however, is *not* a substitute for quick pencil-and-paper circuit analysis, an essential ability that aspiring circuit designers must master. The following series of examples is a step in that direction.

When analyzing a circuit, the first question we need to answer is: *In which mode is the transistor operating*? In some cases, the answer will be obvious. For instance, a quick check of the terminal voltages will indicate whether the transistor is cut off or conducting. If it is conducting, we have to determine whether it is in the active mode or in saturation. In some cases, it won't be obvious; with practice, however, it will become clear in more and more situations. Fortunately, we can always find the answer, using the following procedure.

Assume that the transistor is in the active mode and, using the active-mode model in Table 6.3, determine the various voltages and currents. Then, check for consistency of the results with the assumption of active-mode operation; in other words, is V_{CB} of an *npn* transistor greater than −0.4 V (or V_{CB} of a *pnp* transistor lower than 0.4 V)? If the answer is yes, then our task is complete. If the answer is no, assume saturation-mode operation and, using the saturation-mode model in Table 6.3, determine currents and voltages, and then check whether the results fit with the assumption of saturation-mode operation. Here the test is usually to compute the ratio I_C/I_B and verify that it is lower than the transistor β (i.e.,

$\beta_{forced} < \beta$). For a given transistor type or part number, the manufacturer only gurantees β values will fall within a certain range, say 50 to 150. For this reason, we must use the lowest specified β for this test. Finally, note that the order of these two assumptions can be reversed.

A Note on Units Except when otherwise specified, throughout this book we use a consistent set of units, namely, volts (V), milliamps (mA), and kilohms (kΩ).

Example 6.4

Consider the circuit in Fig. 6.23(a), which we have redrawn in Fig. 6.23(b) to remind you of the convention for indicating connections to dc sources. We want to analyze this circuit to determine all node voltages and branch currents. We will assume that β is specified to be 100.

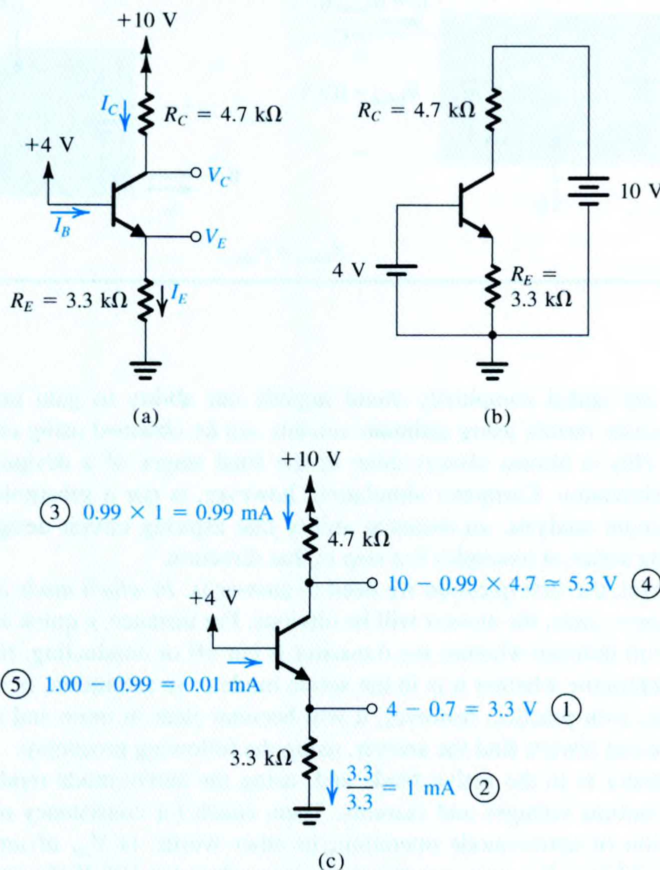


Figure 6.23 Analysis of the circuit for Example 6.4: **(a)** circuit; **(b)** circuit redrawn to remind the reader of the convention used in this book to show connections to the dc sources; **(c)** analysis with the steps numbered.

Solution

Glancing at the circuit in Fig. 6.23(a), we see that the base is connected to +4 V and the emitter is connected to ground through a resistance R_E. Therefore, it is reasonable to conclude that the base–emitter junction is forward biased. Assuming that this is the case and assuming that V_{BE} is approximately 0.7 V, it follows that the emitter voltage is

$$V_E = 4 - V_{BE} \simeq 4 - 0.7 = 3.3 \text{ V}$$

We are now in a good position: because we know the voltages at the two ends of R_E, we can determine the current I_E through it,

$$I_E = \frac{V_E - 0}{R_E} = \frac{3.3 \text{ V}}{3.3 \text{ k}\Omega} = 1 \text{ mA}$$

Since the collector is connected through R_C to the +10-V power supply, it appears possible that the collector voltage will be higher than the base voltage, which implies active-mode operation. Assuming that this is the case, we can evaluate the collector current from

$$I_C = \alpha I_E$$

The value of α is

$$\alpha = \frac{\beta}{\beta + 1} = \frac{100}{101} \simeq 0.99$$

Thus I_C is given by

$$I_C = 0.99 \times 1 = 0.99 \text{ mA}$$

We are now in a position to use Ohm's law to determine the collector voltage V_C,

$$V_C = 10 - I_C R_C = 10 - 0.99 \times 4.7 \simeq +5.3 \text{ V}$$

Since the base is at +4 V, the collector–base junction is reverse biased by 1.3 V, and the transistor is indeed in the active mode as we assumed.

It remains only to determine the base current I_B, as follows:

$$I_B = \frac{I_E}{\beta + 1} = \frac{1}{101} \simeq 0.01 \text{ mA}$$

Before leaving this example, we want to strongly emphasize the value of carrying out the analysis directly on the circuit diagram. Only in this way will one be able to analyze complex circuits in a reasonable length of time. Figure 6.23(c) illustrates our analysis on the circuit diagram, with the order of the analysis steps indicated by the circled numbers.

Example 6.5

We want to analyze the circuit of Fig. 6.24(a) to determine the voltages at all nodes and the currents through all branches. Note that this circuit is identical to that of Fig. 6.23 except that the voltage at the base is now +6 V. Assume that the transistor β is specified to be *at least* 50.

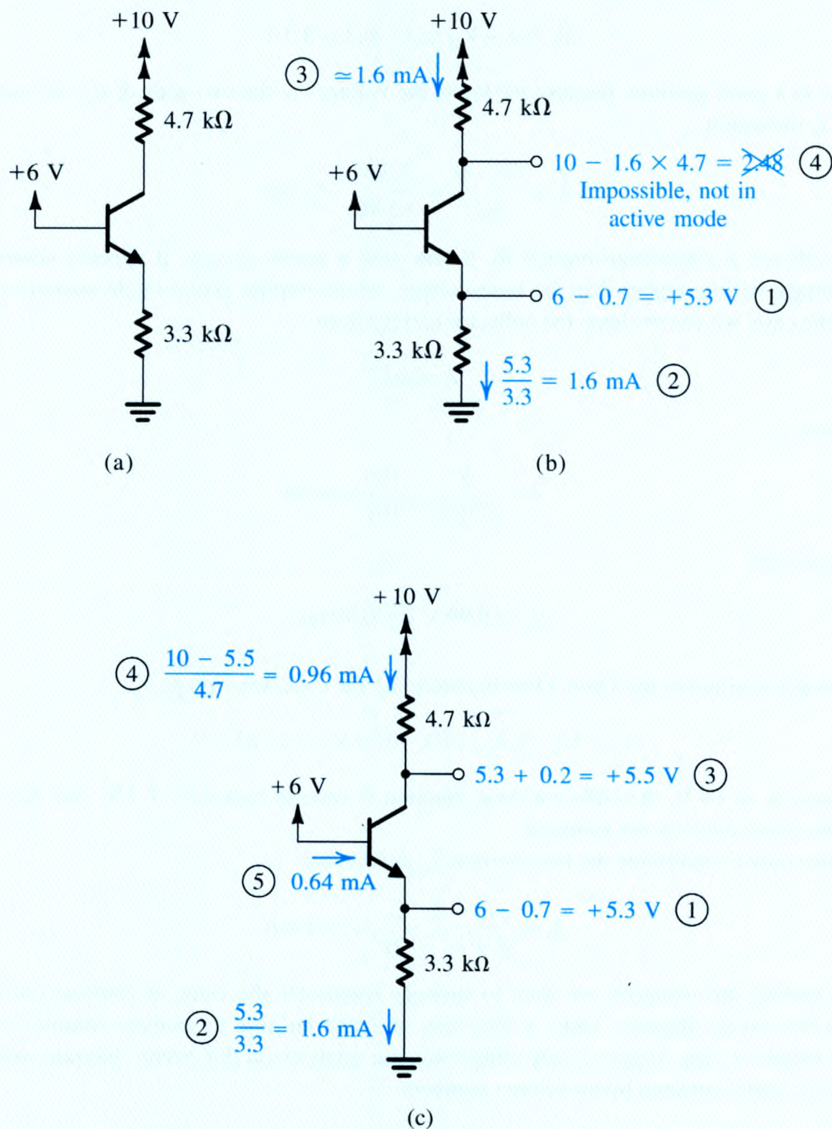


Figure 6.24 Analysis of the circuit for Example 6.5. The circled numbers indicate the order of the analysis steps.

Solution

With +6 V at the base, the base–emitter junction will be forward biased; thus,

$$V_E = +6 - V_{BE} \simeq 6 - 0.7 = 5.3 \text{ V}$$

and

$$I_E = \frac{5.3 \text{ V}}{3.3 \text{ k}\Omega} = 1.6 \text{ mA}$$

Now, assuming active-mode operation, $I_C = \alpha I_E \simeq I_E$; thus,

$$V_C = +10 - 4.7 \times I_C \simeq 10 - 7.52 = 2.48 \text{ V}$$

The analysis is shown in Fig. 6.24(b).

Since the collector voltage calculated is less than the base voltage by 3.52 V, our original assumption of active-mode operation is incorrect: The transistor has to be in the *saturation* mode. Assuming this to be the case, the values of V_E and I_E will remain unchanged. The collector voltage, however, becomes

$$V_C = V_E + V_{CE\text{sat}} \simeq 5.3 + 0.2 = +5.5 \text{ V}$$

from which we can determine I_C as

$$I_C = \frac{10 - 5.5}{4.7} = 0.96 \text{ mA}$$

and I_B can now be found as

$$I_B = I_E - I_C = 1.6 - 0.96 = 0.64 \text{ mA}$$

Thus the transistor is operating at a forced β of

$$\beta_{\text{forced}} = \frac{I_C}{I_B} = \frac{0.96}{0.64} = 1.5$$

The analysis is shown in Fig. 6.24(c). Since β_{forced} is less than the *minimum* specified value of β, the transistor is indeed saturated.

Remember to always use the minimum value of β when testing for saturation. By the same token, when designing a circuit in which a transistor is to be saturated, we should base the design on the minimum specified β. Obviously, if a transistor with this minimum β is saturated, then transistors with higher values of β will also be saturated.

Example 6.6

We want to analyze the circuit in Fig. 6.25(a) to determine the voltages at all nodes and the currents through all branches. Note that this circuit is identical to the one considered in Examples 6.4 and 6.5 except that now the base voltage is zero.

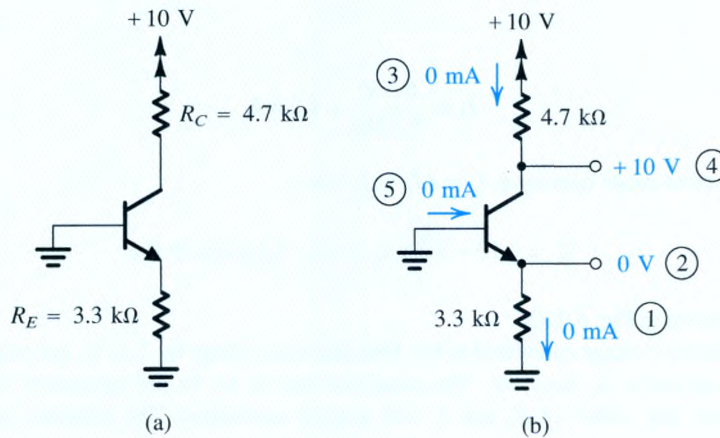


Figure 6.25 Example 6.6: (**a**) circuit; (**b**) analysis, with the order of the analysis steps indicated by circled numbers.

Solution

Since the base is at zero volts and the emitter is connected to ground through R_E, the base–emitter junction cannot conduct and the emitter current is zero. Also, the collector–base junction cannot conduct, since the *n*-type collector is connected through R_C to the positive power supply while the *p*-type base is at ground. It follows that the collector current is zero. The base current also has to be zero, and the transistor is in the *cutoff* mode of operation. The emitter voltage is zero, while the collector voltage is equal to +10 V, since the voltage drops across R_E and R_C are zero. Figure 6.25(b) shows the analysis details.

EXERCISES

D6.22 For the circuit in Fig. 6.23(a), find the highest voltage to which the base can be raised while the transistor remains in the active mode. Assume $\alpha \simeq 1$.
Ans. +4.7 V

D6.23 Redesign the circuit of Fig. 6.23(a) (i.e., find new values for R_E and R_C) to establish a collector current of 0.5 mA and a reverse-bias voltage on the collector–base junction of 2 V. Assume $\alpha \simeq 1$.
Ans. $R_E = 6.6$ kΩ; $R_C = 8$ kΩ

D6.24 For the circuit in Fig. 6.24(a), find the value to which the base voltage should be changed so that the transistor operates in saturation with a forced β of 5.
Ans. +5.18 V

Video Example VE 6.2

The transistor in Fig. VE6.2 has a very high β. Find V_E and V_C for V_B (a) +2.0 V, (b) +1.7 V, and (c) 0 V.

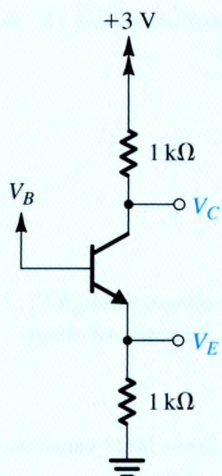


Figure VE6.2 Circuit for Video Example 6.2.

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

VE 6.2

Related end-of-chapter problem: 6.51

Example 6.7

We want to analyze the circuit of Fig. 6.26(a) to determine the voltages at all nodes and the currents through all branches.

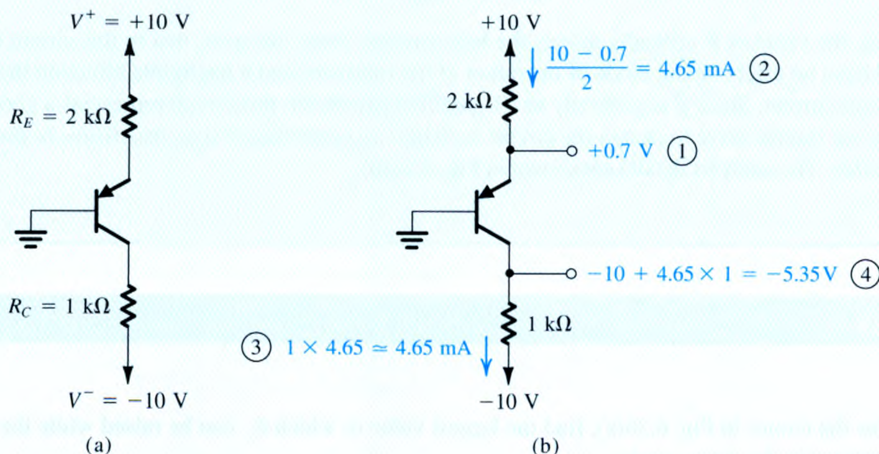


Figure 6.26 Example 6.7: (**a**) circuit; (**b**) analysis, with the steps indicated by circled numbers. The value of β is assumed to be infinite.

Example 6.7 *continued*

Solution

The base of this *pnp* transistor is grounded, while the emitter is connected to a positive supply ($V^+ = +10$ V) through R_E. It follows that the emitter–base junction is forward biased with

$$V_E = V_{EB} \simeq 0.7 \text{ V}$$

Thus the emitter current is given by

$$I_E = \frac{V^+ - V_E}{R_E} = \frac{10 - 0.7}{2} = 4.65 \text{ mA}$$

Since the collector is connected to a negative supply (more negative than the base voltage) through R_C, it is *possible* that this transistor is operating in the active mode. Assuming that this is the case, we obtain

$$I_C = \alpha I_E$$

Since no value for β has been given, we shall assume $\beta = \infty$, which means $\alpha = 1$. Since large variations in β result in small differences in α, this assumption is not critical to determining I_C. Thus,

$$I_C = 1 \times 4.65 = 4.65 \text{ mA}$$

The collector voltage is

$$\begin{aligned} V_C &= V^- + I_C R_C \\ &= -10 + 4.65 \times 1 = -5.35 \text{ V} \end{aligned}$$

Thus the collector–base junction is reverse biased by 5.35 V, and the transistor is indeed in the active mode, which supports our original assumption.

To calculate the base current, we must assume a more realistic value for β, say $\beta = 100$; thus

$$I_B = \frac{I_E}{\beta + 1} = \frac{4.65}{101} \simeq 0.05 \text{ mA}$$

Obviously, the value of β critically affects the base current. Note, however, that in this circuit the value of β will have no effect on the mode of operation of the transistor and a negligible effect on the value of the collector current. Since β is generally an ill-specified parameter, this circuit represents a good design. As a rule, we should strive to *design the circuit such that its performance is as insensitive to the value of β as possible.* The analysis details are shown in Fig. 6.26(b).

EXERCISES

D6.25 For the circuit in Fig. 6.26(a), find the largest value to which R_C can be raised while the transistor remains in the active mode.
Ans. 2.24 kΩ

D6.26 Redesign the circuit of Fig. 6.26(a) (i.e., find new values for R_E and R_C) to establish a collector current of 1 mA and a reverse bias on the collector–base junction of 4 V. Assume $\alpha \simeq 1$.
Ans. $R_E = 9.3\ \text{k}\Omega$; $R_C = 6\ \text{k}\Omega$

Video Example VE 6.3

A single measurement indicates the emitter voltage of the transistor in the circuit of Fig. VE6.3 to be 1.0 V. Under the assumption that $|V_{BE}| = 0.7$ V, what are V_B, I_B, I_E, I_C, V_C, β, and α?

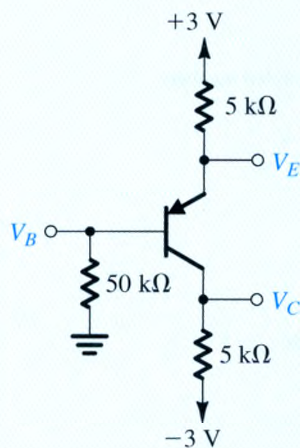


Figure VE6.3 Circuit for Video Example 6.3.

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.


VE 6.3

Related end-of-chapter problem: 6.56

Example 6.8

We want to analyze the circuit in Fig. 6.27(a) to determine the voltages at all nodes and the currents in all branches. Assume $\beta = 100$.

Example 6.8 *continued*

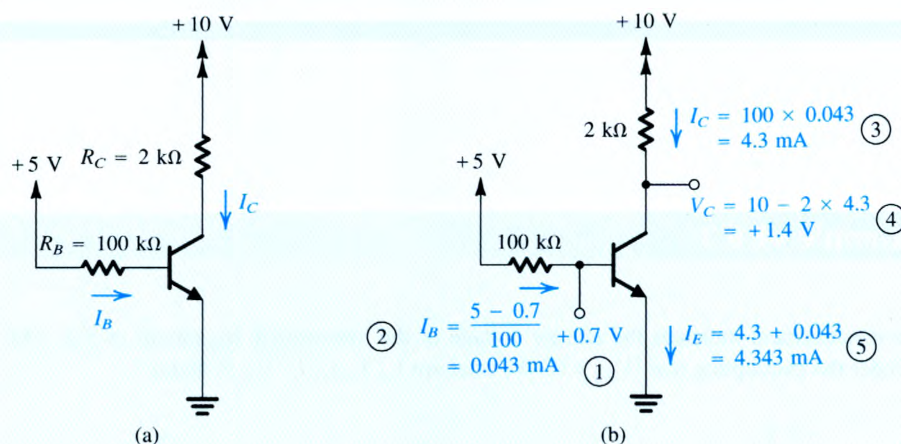


Figure 6.27 Example 6.8: (**a**) circuit; (**b**) analysis, with the steps indicated by the circled numbers.

Solution

The base–emitter junction is clearly forward biased. Thus,

$$I_B = \frac{+5 - V_{BE}}{R_B} \simeq \frac{5 - 0.7}{100} = 0.043 \text{ mA}$$

Assume that the transistor is operating in the active mode. We now can write

$$I_C = \beta I_B = 100 \times 0.043 = 4.3 \text{ mA}$$

The collector voltage can now be determined as

$$V_C = 10 - I_C R_C = 10 - 4.3 \times 2 = +1.4 \text{ V}$$

Since the base voltage V_B is

$$V_B = V_{BE} \simeq +0.7 \text{ V}$$

it follows that the collector–base junction is reverse biased by 0.7 V and the transistor is indeed in the active mode. The emitter current is given by

$$I_E = (\beta + 1) I_B = 101 \times 0.043 = 4.343 \text{ mA}$$

We note from this example that the collector and emitter currents depend critically on the value of β. In fact, if β were 10% higher, the transistor would leave the active mode and enter saturation. Therefore this clearly is a *bad* design. The analysis details are illustrated in Fig. 6.27(b).

EXERCISE

D6.27 The circuit of Fig. 6.27(a) is to be fabricated using a transistor type whose β is specified to be in the range of 50 to 150. That is, individual units of this same transistor type can have β values anywhere in this range. Redesign the circuit by selecting a new value for R_C so that all fabricated circuits are guaranteed to be in the active mode. What is the range of collector voltages that the fabricated circuits may exhibit?

Ans. $R_C = 1.5\ \text{k}\Omega$; $V_C = 0.3$ V to 6.8 V

Example 6.9

We want to analyze the circuit of Fig. 6.28(a) to determine the voltages at all nodes and the currents through all branches. The minimum value of β is specified to be 30.

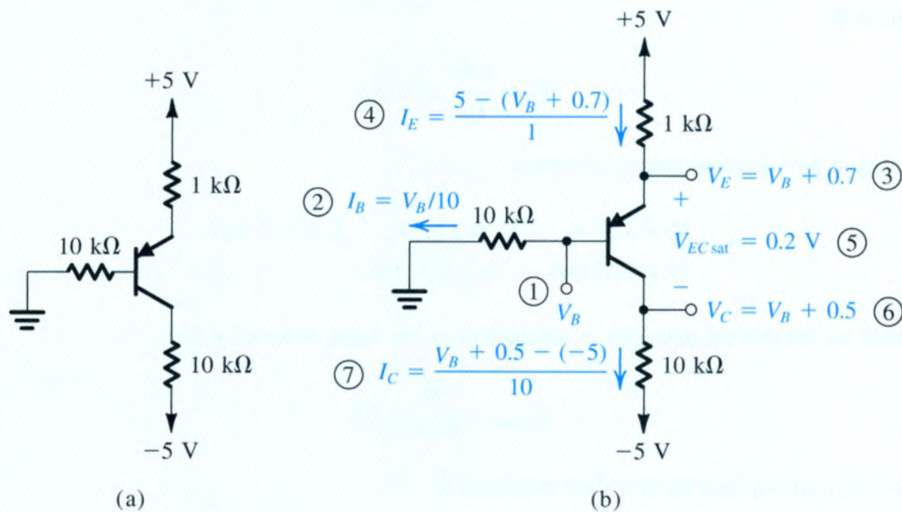


Figure 6.28 Example 6.9: (**a**) circuit; (**b**) analysis with steps numbered.

Solution

A quick glance at this circuit reveals that the transistor is either active or saturated. Assuming active-mode operation and neglecting the base current, we see that the base voltage is approximately zero volts, the emitter voltage is approximately +0.7 V, and the emitter current is approximately 4.3 mA. Since the

Example 6.9 *continued*

maximum current that the collector can support while the transistor remains in the active mode is approximately 0.5 mA, the transistor is definitely saturated.

Assuming that the transistor is saturated and denoting the voltage at the base by V_B (refer to Fig. 6.28b), it follows that

$$V_E = V_B + V_{EB} \simeq V_B + 0.7$$

$$V_C = V_E - V_{ECsat} \simeq V_B + 0.7 - 0.2 = V_B + 0.5$$

$$I_E = \frac{+5 - V_E}{1} = \frac{5 - V_B - 0.7}{1} = 4.3 - V_B \quad \text{mA}$$

$$I_B = \frac{V_B}{10} = 0.1 V_B \quad \text{mA}$$

$$I_C = \frac{V_C - (-5)}{10} = \frac{V_B + 0.5 + 5}{10} = 0.1 V_B + 0.55 \quad \text{mA}$$

Using the relationship $I_E = I_B + I_C$, we obtain

$$4.3 - V_B = 0.1 V_B + 0.1 V_B + 0.55$$

which gives us

$$V_B = \frac{3.75}{1.2} \simeq 3.13 \text{ V}$$

Substituting in the equations above, we obtain

$$V_E = 3.83 \text{ V} \qquad V_C = 3.63 \text{ V} \qquad I_E = 1.17 \text{ mA}$$

$$I_C = 0.86 \text{ mA} \qquad I_B = 0.31 \text{ mA}$$

from which we see that the transistor is saturated since the value of forced β is

$$\beta_{\text{forced}} = \frac{0.86}{0.31} \simeq 2.8$$

which is much smaller than the specified minimum β.

Example 6.10

We want to analyze the circuit of Fig. 6.29(a) to determine the voltages at all nodes and the currents through all branches. Assume $\beta = 100$.

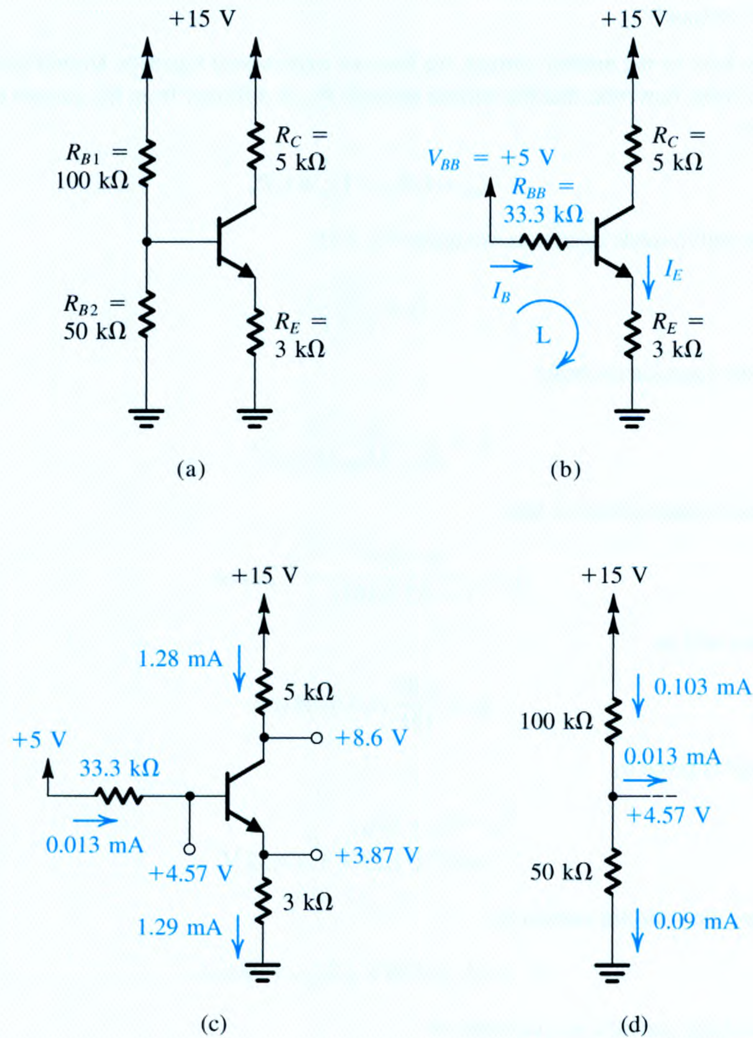


Figure 6.29 Circuits for Example 6.10.

Solution

The first step is to simplify the base circuit using Thévenin's theorem. The result is shown in Fig. 6.29(b), where

$$V_{BB} = +15\frac{R_{B2}}{R_{B1}+R_{B2}} = 15\frac{50}{100+50} = +5\text{ V}$$

$$R_{BB} = R_{B1} \,\|\, R_{B2} = 100 \,\|\, 50 = 33.3\text{ k}\Omega$$

Example 6.10 *continued*

To evaluate the base or the emitter current, we have to write a loop equation around the loop labeled L in Fig. 6.29(b). Note, however, that the current through R_{BB} is different from the current through R_E. The loop equation is

$$V_{BB} = I_B R_{BB} + V_{BE} + I_E R_E$$

Now, assuming active-mode operation, we replace I_B with

$$I_B = \frac{I_E}{\beta + 1}$$

and rearrange the equation to obtain

$$I_E = \frac{V_{BB} - V_{BE}}{R_E + [R_{BB}/(\beta + 1)]}$$

For the numerical values given we have

$$I_E = \frac{5 - 0.7}{3 + (33.3/101)} = 1.29 \text{ mA}$$

The base current will be

$$I_B = \frac{1.29}{101} = 0.0128 \text{ mA}$$

The base voltage is given by

$$V_B = V_{BE} + I_E R_E$$
$$= 0.7 + 1.29 \times 3 = 4.57 \text{ V}$$

We can evaluate the collector current as

$$I_C = \alpha I_E = 0.99 \times 1.29 = 1.28 \text{ mA}$$

The collector voltage can now be evaluated as

$$V_C = +15 - I_C R_C = 15 - 1.28 \times 5 = 8.6 \text{ V}$$

It follows that the collector is higher in potential than the base by 4.03 V, which means that the transistor is in the active mode, as we had assumed. The results of the analysis are given in Fig. 6.29(c, d).

EXERCISE

6.28 If the transistor in the circuit of Fig. 6.29(a) is replaced with another having half the value of β (i.e., $\beta = 50$), find the new value of I_C, and express the change in I_C as a percentage.
Ans. $I_C = 1.15$ mA; –10%

Example 6.11

We want to analyze the circuit in Fig. 6.30(a) to determine the voltages at all nodes and the currents through all branches.

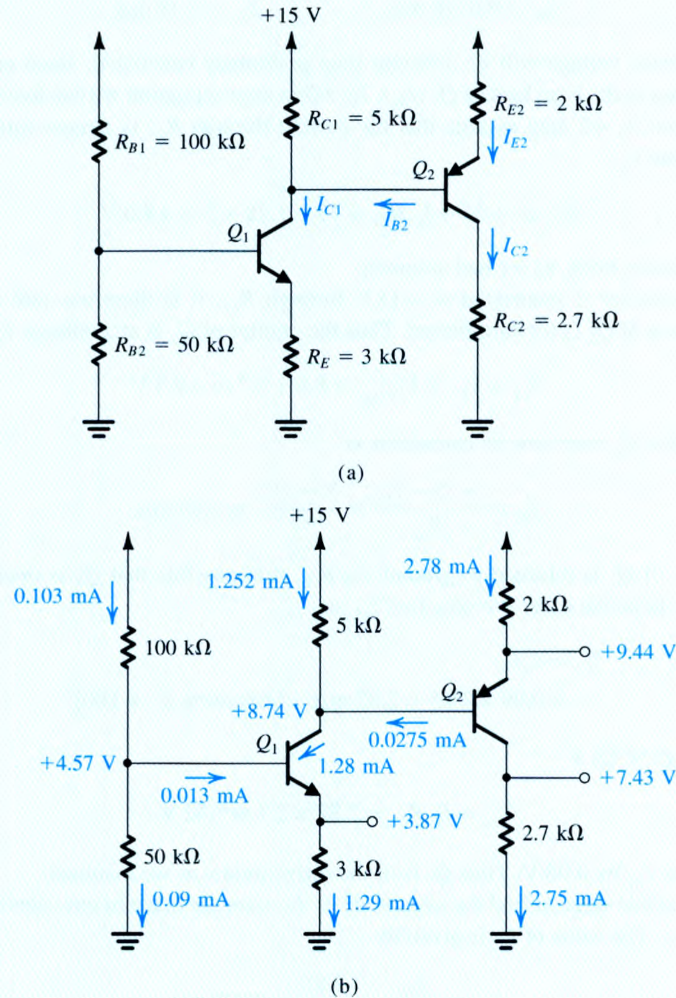


Figure 6.30 Circuits for Example 6.11.

Solution

First, we recognize that part of this circuit is identical to the circuit we analyzed in Example 6.10—namely, the circuit of Fig. 6.29(a). The difference, of course, is that the new circuit has an additional transistor Q_2 together with its associated resistors R_{E2} and R_{C2}. Assume that Q_1 is still in the

Example 6.11 *continued*

active mode. The following values will be identical to those obtained in the previous example:

$$V_{B1} = +4.57 \text{ V} \qquad I_{E1} = 1.29 \text{ mA}$$
$$I_{B1} = 0.0128 \text{ mA} \qquad I_{C1} = 1.28 \text{ mA}$$

However, the collector voltage will be different than previously calculated, since part of the collector current I_{C1} now flows in the base lead of Q_2 (I_{B2}). As a first approximation we can assume that I_{B2} is much smaller than I_{C1}; that is, we may assume that the current through R_{C1} is almost equal to I_{C1}. This will enable us to calculate V_{C1}:

$$V_{C1} \simeq +15 - I_{C1}R_{C1} = 15 - 1.28 \times 5 = +8.6 \text{ V}$$

Thus Q_1 is in the active mode, as we had assumed.

As for Q_2, its emitter is connected to +15 V through R_{E2}. It is therefore safe to assume that the emitter–base junction of Q_2 is forward biased. Thus the emitter of Q_2 is at a voltage V_{E2} given by

$$V_{E2} = V_{C1} + V_{EB}\big|_{Q_2} \simeq 8.6 + 0.7 = +9.3 \text{ V}$$

The emitter current of Q_2 may now be calculated as

$$I_{E2} = \frac{+15 - V_{E2}}{R_{E2}} = \frac{15 - 9.3}{2} = 2.85 \text{ mA}$$

Since the collector of Q_2 is returned to ground via R_{C2}, it is possible that Q_2 is operating in the active mode. Assume this to be the case. We now find I_{C2} as

$$\begin{aligned} I_{C2} &= \alpha_2 I_{E2} \\ &= 0.99 \times 2.85 = 2.82 \text{ mA} \quad (\text{assuming } \beta_2 = 100) \end{aligned}$$

The collector voltage of Q_2 is

$$V_{C2} = I_{C2}R_{C2} = 2.82 \times 2.7 = 7.62 \text{ V}$$

which is lower than V_{B2} by 0.98 V. Thus Q_2 is in the active mode, as we assumed.

It is important at this stage to find the magnitude of the error we made in our calculations by assuming that I_{B2} is negligible. The value of I_{B2} is given by

$$I_{B2} = \frac{I_{E2}}{\beta_2 + 1} = \frac{2.85}{101} = 0.028 \text{ mA}$$

which is indeed much smaller than I_{C1} (1.28 mA). If desired, we can obtain more accurate results by iterating one more time, assuming I_{B2} to be 0.028 mA. The new values will be

$$\text{Current in } R_{C1} = I_{C1} - I_{B2} = 1.28 - 0.028 = 1.252 \text{ mA}$$

$$V_{C1} = 15 - 5 \times 1.252 = 8.74 \text{ V}$$

$$V_{E2} = 8.74 + 0.7 = 9.44 \text{ V}$$

$$I_{E2} = \frac{15 - 9.44}{2} = 2.78 \text{ mA}$$
$$I_{C2} = 0.99 \times 2.78 = 2.75 \text{ mA}$$
$$V_{C2} = 2.75 \times 2.7 = 7.43 \text{ V}$$
$$I_{B2} = \frac{2.78}{101} = 0.0275 \text{ mA}$$

Note that the new value of I_{B2} is very close to the value used in our iteration, and no further iterations are warranted. The final results are indicated in Fig. 6.30(b).

You might be wondering why we need to use an iterative scheme to solve a linear (or linearized) problem. Indeed, we can obtain the exact solution (if we can call anything we are doing with a first-order model exact!) by writing appropriate equations. It is important to emphasize, however, that for most of these problems the approximate solution is quite sufficient, provided we can obtain it quickly and, of course, correctly.

In the above examples, we often used a precise value of α to calculate the collector current. Since $\alpha \simeq 1$, the error in these calculations will be very small if we assume $\alpha = 1$ and $I_C = I_E$. Therefore, except in calculations that depend critically on the value of α (e.g., the calculation of base current), we usually assume $\alpha \simeq 1$.

EXERCISES

6.29 For the circuit in Fig. 6.30, find the total current drawn from the power supply. Hence find the power dissipated in the circuit.
Ans. 4.135 mA; 62 mW

6.30 The circuit in Fig. E6.30 is to be connected to the circuit in Fig. 6.30(a) as indicated; specifically, the base of Q_3 is to be connected to the collector of Q_2. If Q_3 has $\beta = 100$, find the new value of V_{C2} and the values of V_{E3} and I_{C3}.

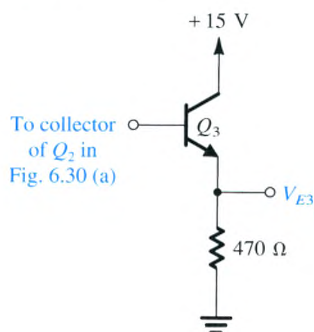


Figure E6.30

Ans. +7.06 V; +6.36 V; 13.4 mA

Example 6.12

We want to evaluate the voltages at all nodes and the currents through all branches in the circuit of Fig. 6.31(a). Assume $\beta = 100$.

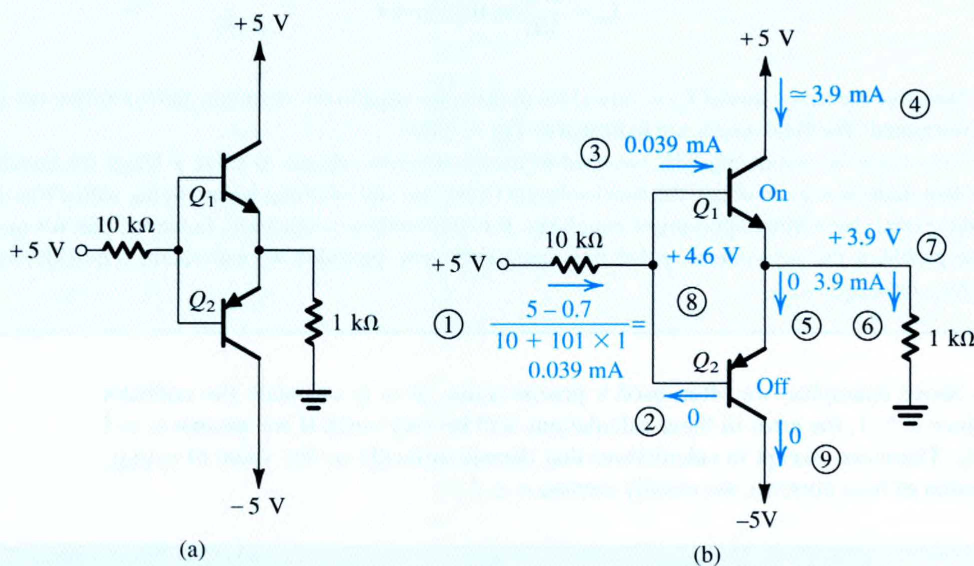


Figure 6.31 Example 6.12: (**a**) circuit; (**b**) analysis with the steps numbered.

Solution

By examining the circuit, we conclude that the two transistors Q_1 and Q_2 cannot be simultaneously conducting: If Q_1 is on, Q_2 will be off, and vice versa. Let's start by assuming Q_2 is on. If it is, then current will flow from ground through the 1-kΩ resistor into the emitter of Q_2. Thus the base of Q_2 will be at a negative voltage, and base current will be flowing out of the base through the 10-kΩ resistor and into the +5-V supply. This is impossible, since if the base is negative, current in the 10-kΩ resistor will have to flow into the base. So we have to conclude that our original assumption—that Q_2 is on—is incorrect. It follows that Q_2 is off and Q_1 is on.

The question now is whether Q_1 is active or saturated. The answer in this case is obvious: Since the base is fed with a +5-V supply and since base current flows into the base of Q_1, it follows that the base of Q_1 will be at a voltage lower than +5 V. Thus the collector–base junction of Q_1 is reverse biased and Q_1 is in the active mode. It remains only to determine the currents and voltages using techniques we have already described in detail. The results are given in Fig. 6.31(b).

EXERCISES

6.31 Solve the problem in Example 6.12 for the case of a voltage of –5 V feeding the bases. What voltage appears at the emitters?
Ans. –3.9 V

6.32 Solve the problem in Example 6.12 with the voltage feeding the bases changed to +10 V. Assume that $\beta_{min} = 30$, and find V_E, V_B, I_{C1}, and I_{C2}.
Ans. +4.8 V; +5.5 V; 4.35 mA; 0

Video Example VE 6.4

For the circuit in Fig. VE6.4, find V_B and V_E for $v_I = 0$ V, +2 V, –2.5 V, and –5 V. The BJTs have $\beta = 50$.

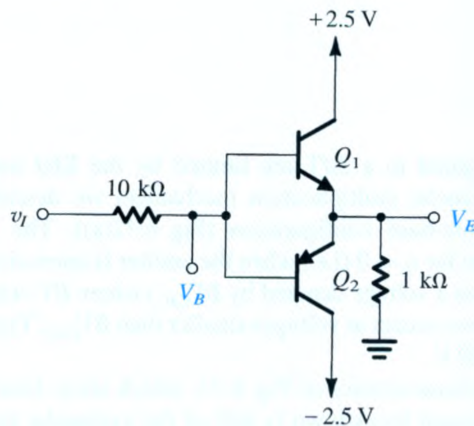


Figure VE6.4 Circuit for Video Example 6.4.

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 6.64

VE 6.4

6.4 Transistor Breakdown and Temperature Effects



We conclude this chapter with a brief discussion of two important nonideal effects in the BJT: voltage breakdown, and the dependence of β on I_C and temperature.

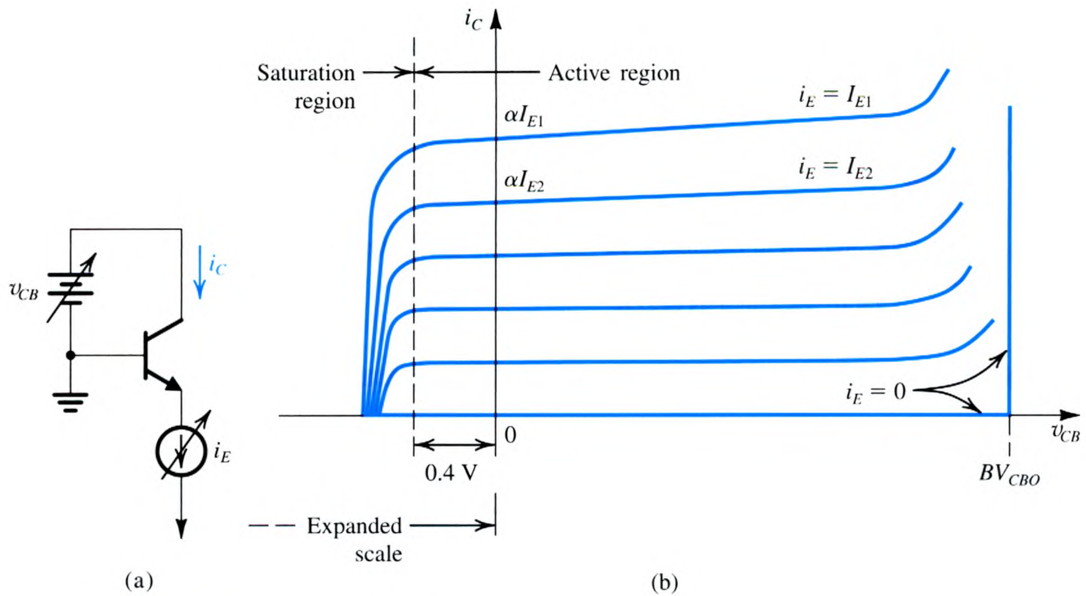


Figure 6.32 The BJT common-base characteristics including the transistor breakdown region.

6.4.1 Transistor Breakdown

The maximum voltages that can be applied to a BJT are limited by the EBJ and CBJ breakdown effects that follow the avalanche multiplication mechanism we described in Section 3.5.3. Consider first the common-base configuration (Fig. 6.32(a)). The $i_C - v_{CB}$ characteristics in Fig. 6.32(b) indicate that for $i_E = 0$ (i.e., when the emitter is open-circuited) the collector–base junction breaks down at a voltage denoted by BV_{CBO} (where BV stands for breakdown voltage). For $i_E > 0$, breakdown occurs at voltages smaller than BV_{CBO}. Typically, for discrete BJTs, BV_{CBO} is greater than 50 V.

Next consider the common-emitter characteristics of Fig. 6.33, which show breakdown occurring at a voltage BV_{CEO}. Here, although breakdown is still of the avalanche type, the effects on the characteristics are more complex than in the common-base configuration. We will not explain these in detail; it is sufficient to point out that typically BV_{CEO} is about half BV_{CBO}. On transistor data sheets, BV_{CEO} is sometimes listed as the **sustaining voltage** LV_{CEO}.

Breakdown of the CBJ in either the common-base or common-emitter configuration is not destructive as long as the power dissipation in the device is kept within safe limits. This, however, is not the case with the breakdown of the emitter–base junction. The EBJ breaks down in an avalanche manner at a voltage BV_{EBO} much smaller than BV_{CBO}. Typically, BV_{EBO} is in the range of 6 V to 8 V, and the breakdown is destructive in the sense that the β of the transistor is permanently reduced. This does not prevent us from using the EBJ as a zener diode to generate reference voltages in IC design, because in that case we aren't concerned with the β-degradation effect. Transistor breakdown and the maximum allowable power dissipation are important parameters in the design of power amplifiers, as we shall see in Chapter 12.

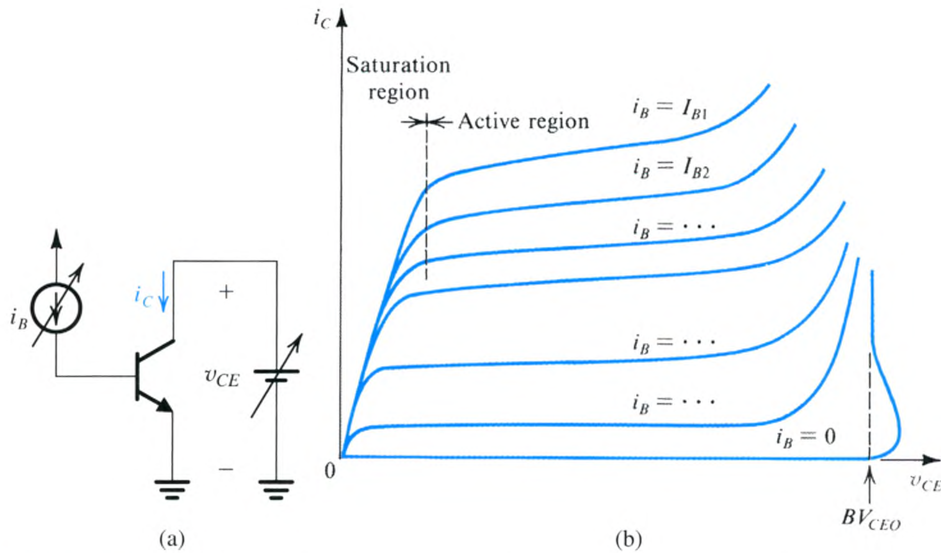


Figure 6.33 The BJT common-emitter characteristics including the breakdown region.

EXERCISE

6.33 What is the output voltage of the circuit in Fig. E6.33 if the transistor $BV_{BCO} = 70$ V?

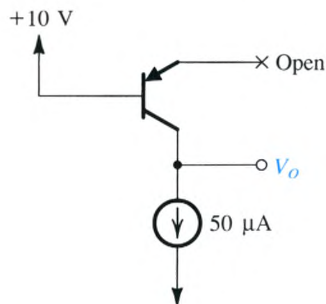


Figure E6.33

Ans. –60 V

6.4.2 Dependence of β on I_C and Temperature

Throughout this chapter we have assumed that the transistor common-emitter dc current gain, β, is constant for a given transistor. In fact, β depends on the dc current at which the transistor is operating, as shown in Fig. 6.34. The physical processes that give rise to this dependence

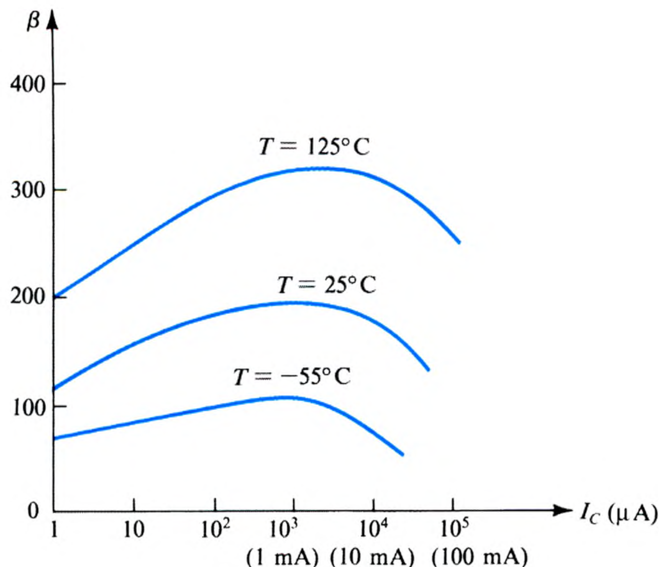


Figure 6.34 Typical dependence of β on I_C and on temperature in an integrated-circuit *npn* silicon transistor intended for operation around 1 mA.

are beyond the scope of this book. Note, however, that there is a current range over which β is highest. Normally, we would arrange to operate the transistor at a current within this range.

Figure 6.34 also shows the dependence of β on temperature. The fact that β increases with temperature can lead to serious problems in transistors that operate at large power levels (see Chapter 12).

Summary

- Depending on the bias conditions on its two junctions, the BJT can operate in one of three possible modes: cutoff (both junctions reverse biased), active (the EBJ forward biased and the CBJ reverse biased), and saturation (both junctions forward biased). Refer to Table 6.1.

- For amplifier applications, the BJT is operated in the active mode. Switching applications make use of the cutoff and saturation modes.

- A BJT operating in the active mode provides a collector current $i_C = I_S e^{|v_{BE}|/V_T}$. The base current $i_B = i_C/\beta$, and the emitter current $i_E = i_C + i_B$. Also, $i_C = \alpha i_E$, and thus $\beta = \alpha/(1-\alpha)$ and $\alpha = \beta/(\beta+1)$. See Table 6.2.

- To ensure operation in the active mode, the collector voltage of an *npn* transistor must be kept higher than approximately 0.4 V below the base voltage. For a *pnp* transistor, the collector voltage must be lower than approximately 0.4 V above the base voltage. Otherwise, the CBJ becomes forward biased, and the transistor enters the saturation region.

- At a constant collector current, the magnitude of the base–emitter voltage decreases by about 2 mV for every 1°C rise in temperature.

- The BJT will be at the edge of saturation when $|v_{CE}|$ is reduced to about 0.3 V. In saturation, $|v_{CE}| \simeq 0.2$ V, and the ratio of i_C to i_B is lower than β (i.e., $\beta_{\text{forced}} < \beta$).

- In the active mode, i_C shows a slight dependence on v_{CE}. This phenomenon, known as the Early effect, is modeled by ascribing a finite (i.e., noninfinite) output resistance to the BJT: $r_o = |V_A|/I'_C$, where V_A is the Early voltage and I'_C is the dc collector current without the Early effect taken into account. In discrete circuits, r_o plays a minor role and can usually be neglected. This is *not* the case, however, in integrated-circuit design (Chapter 8).

- We can greatly simplify the dc analysis of transistor circuits by assuming that $|V_{BE}| \simeq 0.7$ V. Refer to Table 6.3.
- If the BJT is conducting, we can assume it is operating in the active mode and, using the active-mode model, determine all currents and voltages. We can then check the validity of the initial assumption by determining whether the CBJ is reverse biased. If it is, the analysis is complete; otherwise, we have to assume the BJT is operating in saturation and redo the analysis using the saturation-mode model and checking at the end that $I_C < \beta I_B$.

PROBLEMS

Computer Simulations Problems

SIM Problems identified by this icon are intended to demonstrate the value of using SPICE simulation to verify hand analysis and design, and to investigate important issues such as allowable signal swing and amplifier nonlinear distortion. Instructions to assist in setting up simulations for all the indicated problems can be found in the corresponding files on the website. Note that if a particular parameter value is not specified in the problem statement, you are to make a reasonable assumption.

Section 6.1: Device Structure and Physical Operation

6.1 The terminal voltages of various *npn* transistors are measured during operation in their respective circuits with the following results (in volts):

Case	E	B	C	Mode
1	0	0.7	0.7	
2	0	0.8	0.1	
3	−0.7	0	1.0	
4	−0.7	0	−0.6	
5	1.3	2.0	5.0	
6	0	0	5.0	

In this table, 0 indicates the reference terminal to which the black (negative) probe of the voltmeter is connected. For each case, identify the mode of operation of the transistor.

6.2 Two transistors, fabricated with the same technology but having different junction areas, when operated at a base-emitter voltage of 0.75 V, have collector currents of 0.4 mA and 2 mA. Find I_S for each device. What are the relative junction areas?

6.3 In a particular technology, a small BJT operating at $v_{BE} = 30V_T$ conducts a collector current of 100 μA. What is the corresponding saturation current? For a transistor in the same technology but with an emitter junction that is 32 times larger, what is the saturation current? What current will this transistor conduct at $v_{BE} = 30V_T$? What is the base–emitter voltage of the latter transistor at $i_C = 1$ mA? Assume active-mode operation in all cases.

6.4 Two transistors have EBJ areas as follows: $A_{E1} = 100\ \mu\text{m} \times 100\ \mu\text{m}$ and $A_{E2} = 0.4\ \mu\text{m} \times 0.4\ \mu\text{m}$. If the two transistors are operated in the active mode and conduct equal collector currents, what do you expect the difference in their v_{BE} values to be?

6.5 Find the collector currents that you would expect for operation at $v_{BE} = 700$ mV for two transistors for which $I_S = 10^{-13}$ A and $I_S = 10^{-17}$ A. For the transistor with the larger EBJ, what v_{BE} would provide a collector current equal to that provided by the smaller transistor at $v_{BE} = 700$ mV? Assume active-mode operation in all cases.

6.6 In this problem, we contrast two BJT integrated-circuit fabrication technologies: For the "old" technology, a typical *npn* transistor has $I_S = 2 \times 10^{-15}$ A, and for the "new" technology, a typical *npn* transistor has $I_S = 2 \times 10^{-18}$ A.

These typical devices have vastly different junction areas and base widths. For our purpose here we wish to determine the v_{BE} required to establish a collector current of 0.5 mA in each of the two typical devices. Assume active-mode operation.

6.7 Consider an *npn* transistor whose base–emitter drop is 0.76 V at a collector current of 2 mA. What current will it conduct at $v_{BE} = 0.70$ V? What is its base–emitter voltage for $i_C = 4\ \mu$A?

6.8 In a particular BJT operating in the active mode, the base current is 10 μA, and the collector current is 900 μA. Find β and α for this device.

6.9 Find the values of β that correspond to α values of 0.5, 0.8, 0.9, 0.95, 0.98, 0.99, 0.995, and 0.999.

6.10 Find the values of α that correspond to β values of 1, 2, 10, 20, 50, 100, 200, 500, and 1000.

***6.11** For a transistor with α close to unity, show that if α changes by a small per-unit amount $(\Delta\alpha/\alpha)$, the corresponding per-unit change in β is given approximately by

$$\frac{\Delta\beta}{\beta} \simeq \beta\left(\frac{\Delta\alpha}{\alpha}\right)$$

Now, for a transistor whose nominal β is 100, find the percentage change in its α value corresponding to a drop in its β of 10%.

6.12 An *npn* transistor of a type whose β is specified to range from 50 to 200 is connected in a circuit with emitter grounded, collector at + 10 V, and a current of 10 μA injected into the base. Calculate the range of collector and emitter currents that can result. What is the maximum power dissipated in the transistor? (*Note:* Perhaps you can see why this is a bad way to establish the operating current in the collector of a BJT.)

6.13 A BJT is specified to have $I_S = 2 \times 10^{-15}$ A and β that falls in the range of 50 to 200. If the transistor is operated in the active mode with v_{BE} set to 0.700 V, find the expected range of i_C, i_B, and i_E.

6.14 Measurements made on a number of transistors operating in the active mode with $i_E = 1$ mA indicate base currents of 10 μA, 20 μA, and 50 μA. For each device, find i_C, β, and α.

6.15 Measurements of V_{BE} and two terminal currents taken on a number of *npn* transistors operating in the active mode are in the following table. For each, calculate the missing current value as well as α, β, and I_S as indicated by the table.

Transistor	a	b	c	d	e
V_{BE} (mV)	700	690	580	780	820
I_C (mA)	1.000	1.000		10.10	
I_B (μA)	10		5	120	1050
I_E (mA)		1.020	0.235		75.00
α					
β					
I_S					

6.16 When operated in the active mode, a particular *npn* BJT conducts a collector current of 1 mA and has $v_{BE} = 0.7$ V and $i_B = 10\ \mu$A. Use these data to create specific transistor models of the form shown in Figs. 6.5(a) to (d).

6.17 Using the *npn* transistor model of Fig. 6.5(b), consider the case of a transistor for which the base is connected to ground, the collector is connected to a 3-V dc source through a 2-kΩ resistor, and a 1-mA current source is connected to the emitter with the polarity so that current is drawn out of the emitter terminal. If $\beta = 100$ and $I_S = 5 \times 10^{-15}$ A, find the voltages at the emitter and the collector and calculate the base current.

D 6.18 Consider an *npn* transistor operated in the active mode and represented by the model of Fig. 6.5(d). Let the transistor be connected as indicated by the equivalent circuit shown in Fig. 6.6(b). Calculate the values of R_B and R_C that will establish a collector current I_C of 0.5 mA and a collector-to-emitter voltage V_{CE} of 1 V. The BJT is specified to have $\beta = 50$ and $I_S = 5 \times 10^{-15}$ A, and $V_{CC} = 5$ V.

6.19 An *npn* transistor has a CBJ with an area 100 times that of the EBJ. If $I_S = 10^{-15}$ A, find the voltage drop across EBJ and across CBJ when each is forward biased and conducting a current of 1 mA. Also find the forward current each junction would conduct when forward biased with 0.5 V.

*6.20 In order to investigate the operation of the *npn* transistor in saturation using the model of Fig. 6.9, let $I_S = 10^{-15}$ A, $v_{BE} = 0.7$ V, $\beta = 100$, and $I_{SC}/I_S = 100$. For each of three values of v_{CE} (namely, 0.4 V, 0.3 V, and 0.2 V), find v_{BC}, i_{BC}, i_{BE}, i_B, i_C, and i_C/i_B. Present your results in a table. Also find v_{CE} that results in $i_C = 0$.

*6.21 Use Eqs. (6.14), (6.15), and (6.16) to show that an *npn* transistor operated in saturation exhibits a collector-to-emitter voltage, V_{CEsat}, given by

$$V_{CEsat} = V_T \ln\left[\left(\frac{I_{SC}}{I_S}\right)\frac{1+\beta_{\text{forced}}}{1-\beta_{\text{forced}}/\beta}\right]$$

Use this relationship to evaluate V_{CEsat} for $\beta_{\text{forced}} = 50$, 10, 5, 1, and 0 for a transistor with $\beta = 100$ and with a CBJ area 100 times that of the EBJ. Present your results in a table.

6.22 Consider the *pnp* large-signal model of Fig. 6.11(b) applied to a transistor having $I_S = 10^{-15}$ A and $\beta = 40$. If the emitter is connected to ground, the base is connected to a current source that pulls 10 μA out of the base terminal, and the collector is connected to a negative supply of −5 V via a 10-kΩ resistor, find the collector voltage, the emitter current, and the base voltage.

6.23 A *pnp* transistor has $v_{EB} = 0.6$ V at a collector current of 0.1 mA. What do you expect v_{EB} to become at $i_C = 1$ mA? At $i_C = 10$ mA?

6.24 A *pnp* transistor modeled with the circuit in Fig. 6.11 (b) is connected with its base at ground, collector at −1.0 V, and a 1-mA current is injected into its emitter. If the transistor has $\beta = 20$, what are its base and collector currents? In which direction do they flow? If $I_S = 10^{-15}$ A, what voltage results at the emitter? What does the collector current become if a transistor with $\beta = 200$ is substituted? (*Note:* The fact that the collector current changes by less than 5% for a ten-fold increase in β illustrates that this is a good way to establish a specific collector current.)

6.25 A *pnp* power transistor operates with an emitter-to-collector voltage of 5 V, an emitter current of 5 A, and $V_{EB} = 0.8$ V. For $\beta = 20$, what base current is required? What is I_S for this transistor? Compare the emitter–base junction area of this transistor with that of a small-signal transistor that conducts $i_C = 1$ mA with $v_{EB} = 0.70$ V. How much larger is it?

6.26 While Fig. 6.5 provides four possible large-signal equivalent circuits for the *npn* transistor, only two equivalent circuits for the *pnp* transistor are provided in Fig. 6.11. Supply the missing two.

6.27 By analogy to the *npn* case shown in Fig. 6.9, give the equivalent circuit of a *pnp* transistor in saturation.

Section 6.2: Current–Voltage Characteristics

6.28 For the circuits in Fig. P6.28, assume that the transistors have very large β. Some measurements have been made on these circuits, with the results indicated in the figure. Find the values of the other labeled voltages and currents.

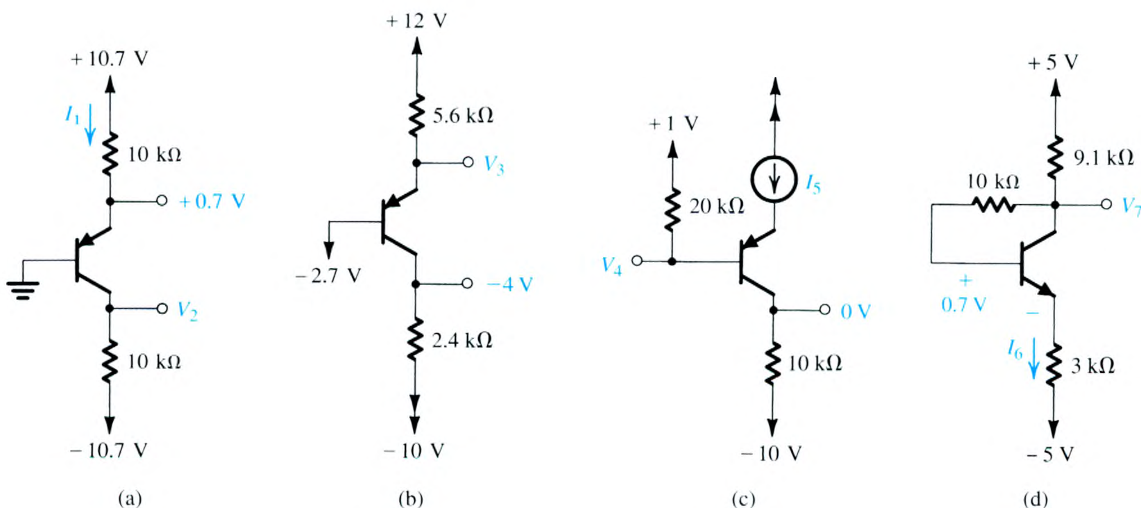


Figure P6.28

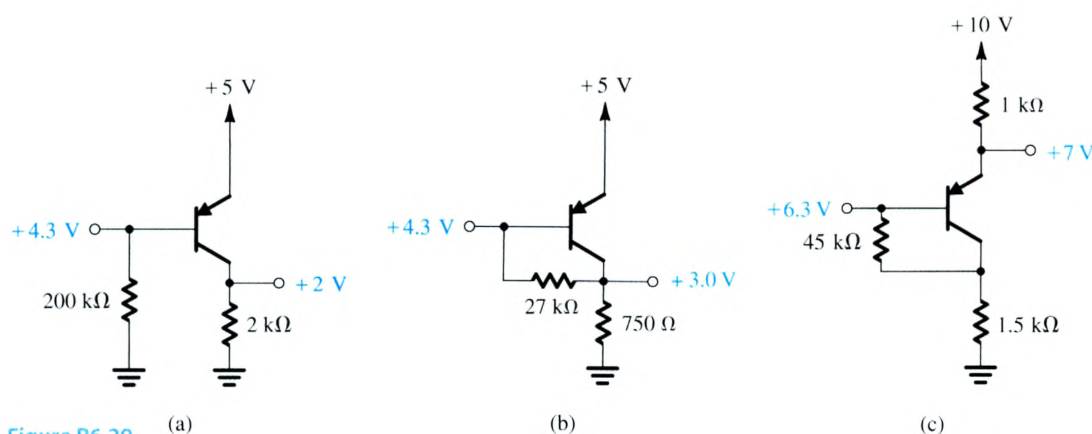


Figure P6.29

6.29 Measurements on the circuits of Fig. P6.29 produce labeled voltages as indicated. Find the value of β for each transistor.

6.30 A very simple circuit for measuring β of an *npn* transistor is shown in Fig. P6.30. In a particular design, V_{CC} is provided by a 3-V battery; M is a current meter with a 50-μA full scale and relatively low resistance that you can neglect for our purposes here. Assuming that the transistor has $V_{BE} = 0.7$ V at $I_E = 1$ mA, what value of R_C would establish a resistor current of 1 mA? Now, to what value of β does a meter reading of full scale correspond? What is β if the meter reading is 1/5 of full scale? 1/10 of full scale?

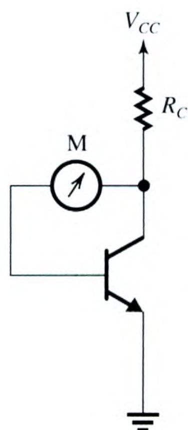


Figure P6.30

6.31 Repeat Exercise 6.13 for the situation in which the power supplies are reduced to ±3 V.

D 6.32 Design the circuit in Fig. P6.32 to establish a current of 0.5 mA in the emitter and a voltage of −0.5 V at the collector. The transistor $v_{EB} = 0.7$ V at $I_E = 1$ mA, and $\beta = 100$. To what value can R_C be increased while the collector current remains unchanged?

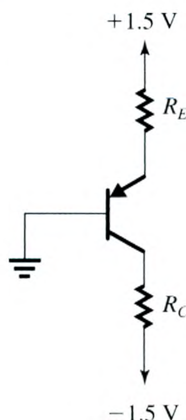


Figure P6.32

D 6.33 The table of standard values for resistors with 5% tolerance in Appendix J shows that the closest values to those found in the design of Example 6.2 are 5.1 kΩ and 6.8 kΩ. For these values, use approximate calculations (e.g., $V_{BE} \simeq 0.7$ V and $\alpha \simeq 1$) to determine the values of collector current and collector voltage that are likely to result.


VE 6.1

D 6.34 Design the circuit in Fig. VE6.1 to establish $I_C = 0.5$ mA and $V_C = 0$ V. The transistor exhibits v_{BE} of 0.7 V at $i_C = 1$ mA, and $\beta = 100$.

6.35 For each of the circuits shown in Fig. P6.35, find the emitter, base, and collector voltages and currents. Use $\beta = 50$, but assume $|V_{BE}| = 0.7$ V independent of current level.

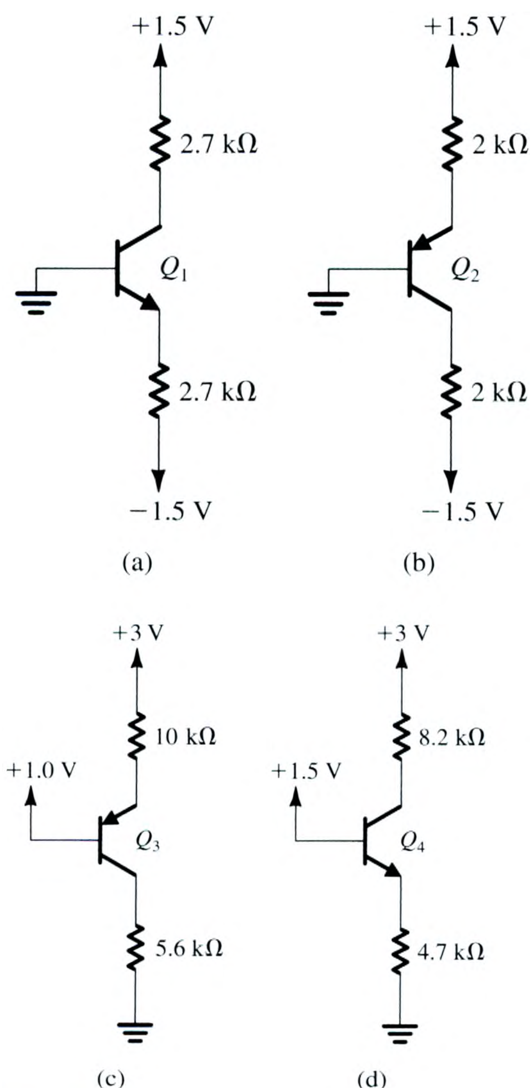


Figure P6.35

6.36 The current I_{CBO} of a small transistor is measured to be 10 nA at 25°C. If the temperature of the device is raised to 125°C, what do you expect I_{CBO} to become?

6.37 Augment the model of the *npn* BJT shown in Fig. 6.19(a) by a current source representing I_{CBO}. Assume that r_o is very large and thus can be neglected. In terms of this addition, what do the terminal currents i_B, i_C, and i_E become? If the base lead is open-circuited while the emitter is connected to ground, and the collector is connected to a positive supply, find the emitter and collector currents.

6.38 A BJT whose emitter current is fixed at 1 mA has a base–emitter voltage of 0.70 V at 25°C. What base–emitter voltage would you expect at 0°C? At 100°C?

6.39 A particular *pnp* transistor operating at an emitter current of 0.5 mA at 20°C has an emitter–base voltage of 692 mV.

(a) What does v_{EB} become if the junction temperature rises to 50°C?
(b) If the transistor is operated at a fixed emitter–base voltage of 700 mV, what emitter current flows at 20°C? At 50°C?

6.40 Consider a transistor for which the base–emitter voltage drop is 0.7 V at 10 mA. What current flows for $v_{BE} = 0.5$ V? Evaluate the ratio of the slopes of the i_C–v_{BE} curve at $v_{BE} = 700$ mV and at $v_{BE} = 500$ mV. The large ratio confirms the point that the BJT has an "apparent threshold" at $v_{BE} \simeq 0.5$ V.

6.41 Use Eq. (6.18) to plot i_C versus v_{CE} for an *npn* transistor having $I_S = 10^{-15}$ A and $V_A = 100$ V. Provide curves for $v_{BE} = 0.65$, 0.70, 0.72, 0.73, and 0.74 volts. Show the characteristics for v_{CE} up to 15 V.

***6.42** In the circuit shown in Fig. P6.42, current source I is 1.1 mA, and at 25°C $v_{BE} = 680$ mV at $i_E = 1$ mA. At 25°C with $\beta = 100$, what currents flow in R_1 and R_2? What voltage would you expect at node E? Noting that the temperature coefficient of v_{BE} for I_E constant is −2 mV/°C, what is the TC of v_E? For an ambient temperature of 75°C, what voltage would you expect at node E? Clearly state any simplifying assumptions you make.

6.43 For a particular *npn* transistor operating at a v_{BE} of 680 mV and $I_C = 0.5$ mA, the i_C–v_{CE} characteristic has a slope of 0.5×10^{-5} ℧. To what value of output resistance does this correspond? What is the value of the Early voltage

for this transistor? For operation at 5 mA, what would the output resistance become?

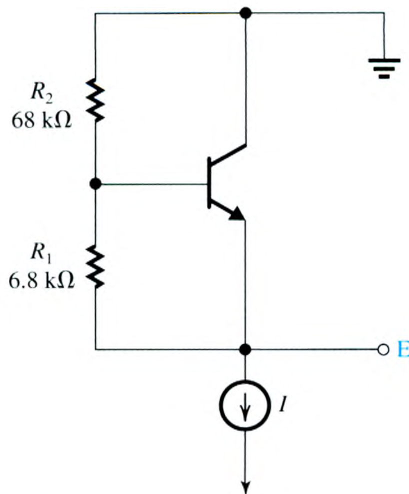


Figure P6.42

6.44 For a BJT having an Early voltage of 40 V, what is its output resistance at 1 mA? At 100 μA?

6.45 Measurements of the i_C–v_{CE} characteristic of a small-signal transistor operating at $v_{BE} = 710$ mV show that $i_C = 1.1$ mA at $v_{CE} = 5$ V and that $i_C = 1.3$ mA at $v_{CE} = 15$ V. What is the corresponding value of i_C near saturation? At what value of v_{CE} is $i_C = 1.2$ mA? What is the value of the Early voltage for this transistor? What is the output resistance that corresponds to operation at $v_{BE} = 710$ mV?

6.46 Give the *pnp* equivalent circuit models that correspond to those shown in Fig. 6.19 for the *npn* case.

6.47 A BJT operating at $i_B = 10\ \mu$A and $i_C = 1.0$ mA undergoes a reduction in base current of 1.0 μA. When v_{CE} is held constant, the corresponding reduction in collector current is 0.08 mA. What are the values of β and the incremental β or β_{ac} that apply? If the base current is increased from 10 μA to 12 μA and v_{CE} is increased from 8 V to 10 V, what collector current results? Assume $V_A = 100$ V.

6.48 For the circuit in Fig. P6.48 let $V_{CC} = 5$ V, $R_C = 10$ kΩ, and $R_B = 100$ kΩ. The BJT has $\beta = 50$. Find the value of V_{BB} that results in the transistor operating

(a) in the active mode with $V_C = 2$ V;
(b) at the edge of saturation;
(c) deep in saturation with $\beta_{\text{forced}} = 10$.

Assume $V_{BE} \simeq 0.7$ V.

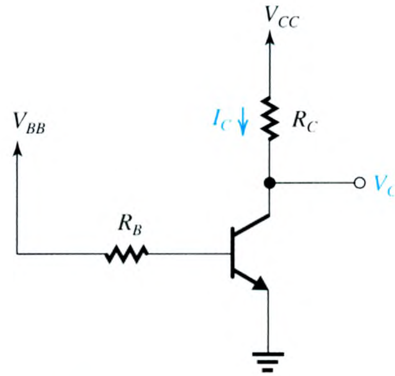


Figure P6.48

SIM **D *6.49** Consider the circuit of Fig. P6.48 for the case $V_{BB} = V_{CC}$. If the BJT is saturated, use the equivalent circuit of Fig. 6.21 to derive an expression for β_{forced} in terms of V_{CC} and (R_B/R_C). Also derive an expression for the total power dissipated in the circuit. For $V_{CC} = 5$ V, design the circuit so that it operates at a forced β as close to 10 as possible while limiting the power dissipation to no larger than 20 mW. Use 1% resistors (see Appendix J).

6.50 The *pnp* transistor in the circuit in Fig. P6.50 has $\beta = 50$. Show that the BJT is operating in the saturation mode and find β_{forced} and V_C. To what value should R_B be increased in order for the transistor to operate at the edge of saturation?

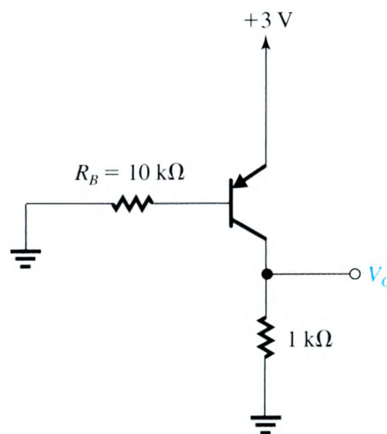


Figure P6.50

Section 6.3: BJT Circuits at DC



VE 6.2

6.51 The transistor in the circuit of Fig. P6.51 has a very high β. Find V_E and V_C for V_B (a) +2.0 V, (b) +1.7 V, and (c) 0 V.

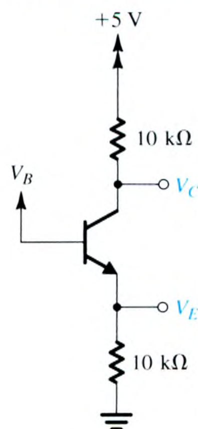


Figure P6.51

6.52 The transistor in the circuit of Fig. P6.51 has a very high β. Find the highest value of V_B for which the transistor still operates in the active mode. Also, find the value of V_B for which the transistor operates in saturation with a forced β of 3.

***6.53** Consider the operation of the circuit shown in Fig. P6.53 for V_B at –1 V, 0 V, and +1 V. Assume that β is very high. What values of V_E and V_C result? At what value of V_B does the emitter current reduce to one-tenth of its value for $V_B = 0$ V? For what value of V_B is the transistor just at the edge of conduction? ($v_{BE} = 0.5$ V) What values of V_E and V_C correspond? For what value of V_B does the transistor reach the edge of saturation? What values of V_C and V_E correspond? Find the value of V_B for which the transistor operates in saturation with a forced β of 1.5.

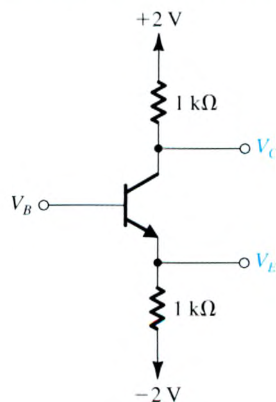


Figure P6.53

6.54 For the transistor shown in Fig. P6.54, assume $\alpha \simeq 1$ and $v_{BE} = 0.5$ V at the edge of conduction. What are the values of V_E and V_C for $V_B = 0$ V? For what value of V_B does the transistor cut off? Saturate? In each case, what values of V_E and V_C result?

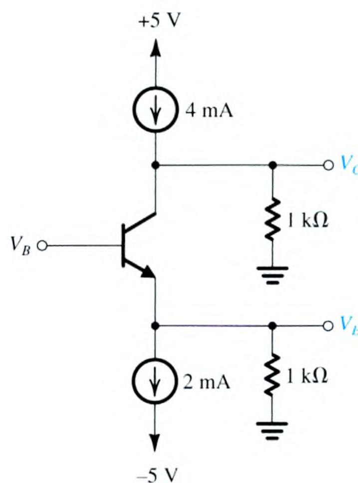


Figure P6.54

D 6.55 Consider the circuit in Fig. P6.51 with the base voltage V_B obtained using a voltage divider across the 5-V supply. Assuming the transistor β to be very large (i.e., ignoring the base current), design the voltage divider to obtain $V_B = 1.5$ V. Design for a 0.1-mA current in the voltage divider. Now, if the BJT $\beta = 100$, analyze the circuit to determine the collector current and the collector voltage.



VE 6.3

6.56 A single measurement indicates the emitter voltage of the transistor in the circuit of Fig. P6.56 to be 1.0 V. Under the assumption that $|V_{BE}| = 0.7$ V, what are V_B, I_B, I_E, I_C, V_C, β, and α ?

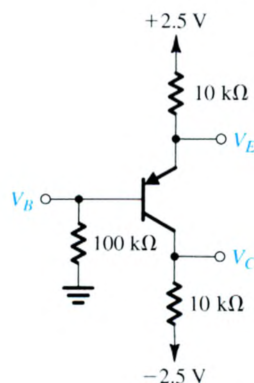


Figure P6.56

D 6.57 Design a circuit using a *pnp* transistor for which $\alpha \simeq 1$ using two resistors connected appropriately to ± 2 V so that $V_B = 0$ V, $I_E = 0.5$ mA, and $V_{BC} = 1$ V. What exact values of R_E and R_C would be needed? Now, consult the table of standard 5% resistor values provided in Appendix J to select suitable practical values. What values of resistors have you chosen? What are the values of I_E and V_{BC} that result?

6.58 In the circuit shown in Fig. P6.58, the transistor has $\beta = 40$. Find the values of V_B, V_E, and V_C. If R_B is raised to 100 kΩ, what voltages result? With $R_B = 100$ kΩ, what value of β would return the voltages to the values first calculated?

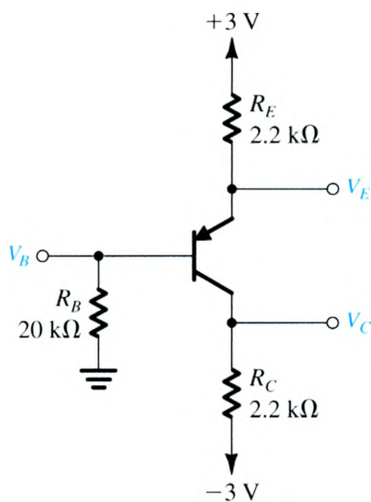


Figure P6.58

6.59 In the circuit shown in Fig. P6.58, the transistor has $\beta = 50$. Find the values of V_B, V_E, and V_C, and verify that the transistor is operating in the active mode. What is the largest value that R_C can have while the transistor remains in the active mode?

SIM **6.60** For the circuit in Fig. P6.60, find V_B, V_E, and V_C for $R_B = 100$ kΩ, 10 kΩ, and 1 kΩ. Let $\beta = 100$.

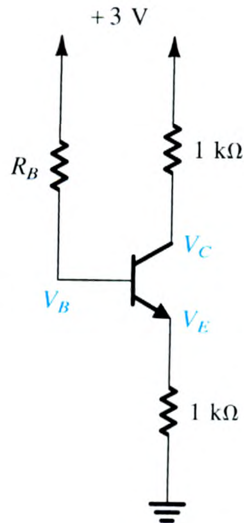


Figure P6.60

6.61 For the circuits in Fig. P6.61, find values for the labeled node voltages and branch currents. Assume β to be very high.

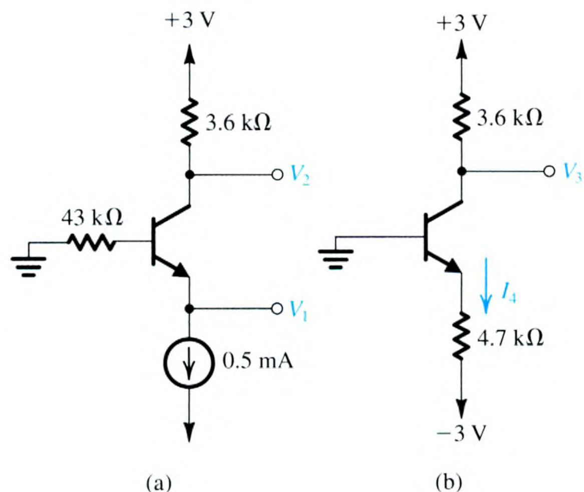


Figure P6.61

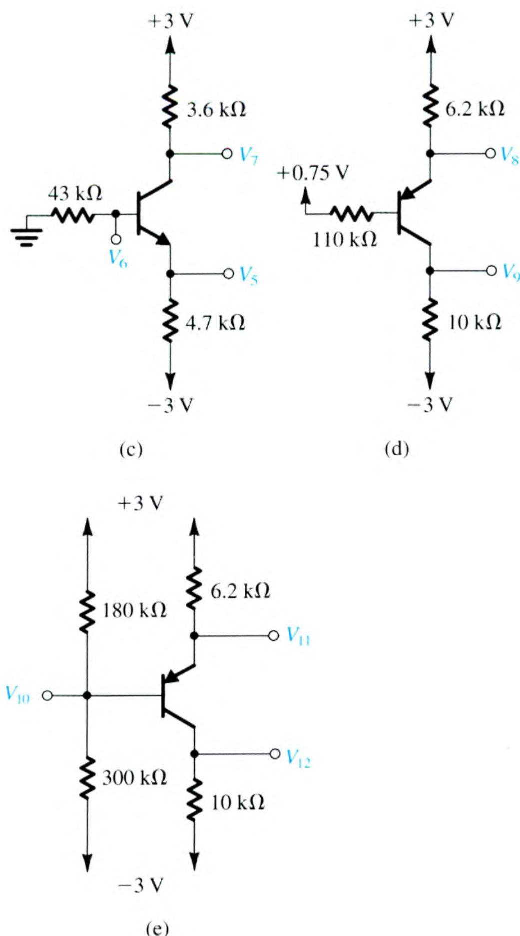


Figure P6.61 *continued*

***6.62** Repeat the analysis of the circuits in Problem 6.61 using $\beta = 100$. Find all the labeled node voltages and branch currents.

D *6.63 Design the circuit in Fig. P6.63 so that a current of 1 mA is established in the emitter and a voltage of −1 V appears at the collector. The transistor used has a nominal β of 100. However, the β value can be as low as 50 and as high as 150. Your design should ensure that the specified emitter current is obtained when $\beta = 100$ and that at the extreme values of β the emitter current does not change by more than 10% of its nominal value. Also, design for as large a value for R_B as possible. Give the values of R_B, R_E, and R_C to the nearest kilohm. What is the expected range of collector current and collector voltage corresponding to the full range of β values?

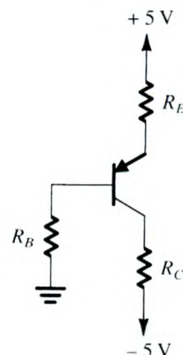


Figure P6.63

D 6.64 The *pnp* transistor in Fig. P6.64 has $\beta = 50$. Find the value for R_C that gives $V_C = +2$ V. What happens if the transistor is replaced with another having $\beta = 100$? Give the value of V_C in the latter case.

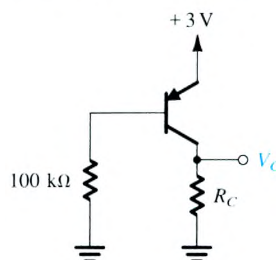


Figure P6.64

6.65 Design the circuit in Fig. P6.65 to obtain $I_E = 0.2$ mA, $V_E = +2$ V, and $V_C = +5$ V. Design for $I_{B2} = 0.1$ mA. Use standard 5% resistors (refer to Table J.1 in Appendix J). The transistor has $V_{BE} = 0.7$ V and $\beta = 100$.

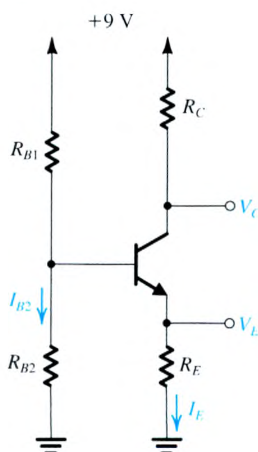


Figure P6.65

***6.66** For the circuit shown in Fig. P6.66, find the labeled node voltages for:

(a) $\beta = \infty$
(b) $\beta = 100$

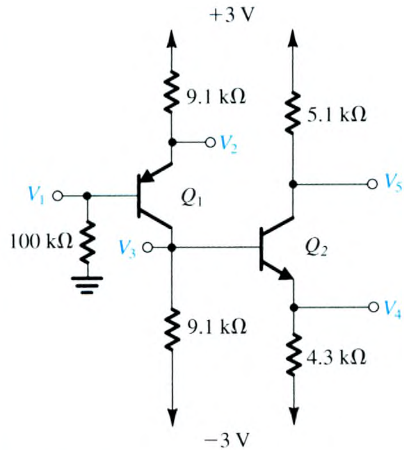


Figure P6.66

D *6.67 Using $\beta = \infty$, design the circuit shown in Fig. P6.67 so that the emitter currents of Q_1, Q_2, and Q_3 are 0.5 mA, 0.5 mA, and 1 mA, respectively, and $V_3 = 0$, $V_5 = -2$ V, and $V_7 = 1$ V. For each resistor, select the nearest standard value utilizing the table of standard values for 5% resistors in Appendix J. Now, for $\beta = 100$, find the values of V_3, V_4, V_5, V_6, and V_7.

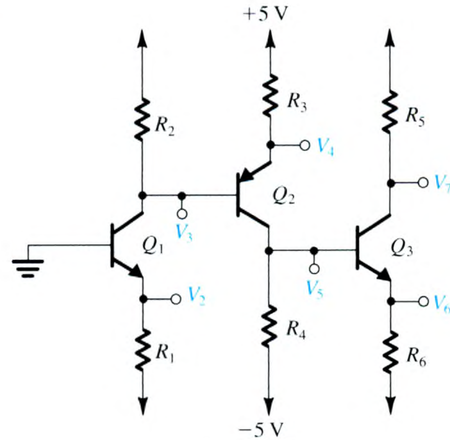


Figure P6.67



6.68 For the circuit in Fig. VE 6.4, find V_B and V_E for $v_I = 0$ V, −2 V, +2.5 V, and +5 V. The BJTs have $\beta = 50$.

****6.69** All the transistors in the circuits of Fig. P6.69 are specified to have a minimum β of 50. Find approximate values for the collector voltages and calculate forced β for each of the transistors. (*Hint:* Begin by assuming that all transistors are operating in saturation, and then verify the assumption.)

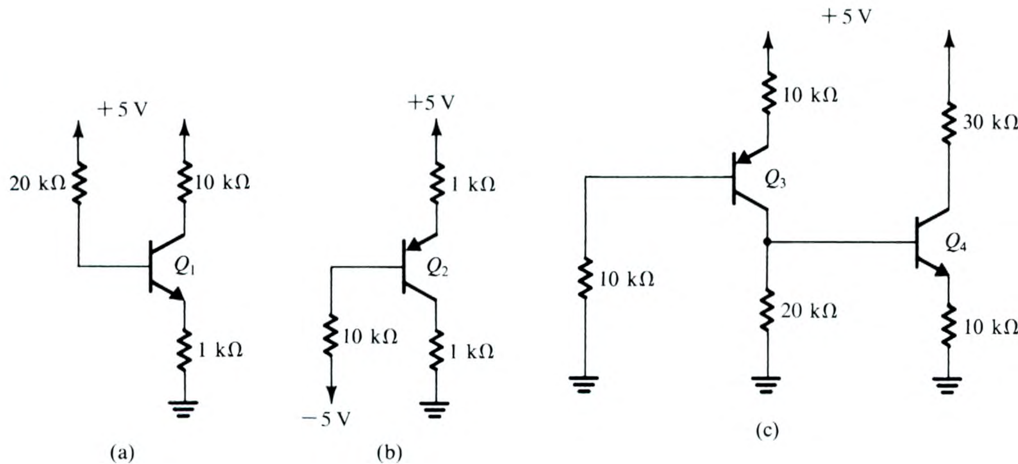


Figure P6.69

CHAPTER 7

Transistor Amplifiers

IN THIS CHAPTER YOU WILL LEARN

- How the transistor (a MOSFET or a BJT) can be used to make an amplifier.
- How to obtain linear amplification from the fundamentally nonlinear MOS and bipolar transistor.
- How to model the linear operation of a transistor around a bias point by an equivalent circuit that can be used in the analysis and design of transistor amplifiers.
- The three basic ways to connect a MOSFET or a BJT to construct amplifiers with different properties.
- Practical circuits for MOS and bipolar transistor amplifiers that can be constructed using discrete components.

Introduction

Having studied the two major transistor types, the MOSFET (Chapter 5) and the BJT (Chapter 6), we now begin to study their application. There are two distinctly different kinds of transistor applications: as a switch, in the design of digital circuits (Chapters 16–18) and as a controlled source, in the design of amplifiers for analog circuits. This chapter and the next six focus on the use of the transistor in the design of various amplifier types.

Since the basic principles that underlie the use of the MOSFET and the BJT in amplifier design are the same, we study the two devices together in this chapter. Besides providing some economy in presentation, this unified study allows us to make important comparisons between MOS and bipolar amplifiers.

The bulk of this chapter is concerned with the fundamental principles and concepts that are the basis for the application of transistors in amplifier design: We study in detail the models that are used to represent both transistor types in the analysis and design of small-signal linear amplifiers. We also study the three basic configurations in which each of the two transistor types can be connected to create an amplifier.

The chapter concludes with examples of discrete-circuit amplifiers. These are circuits that can be assembled using discrete transistors, resistors, and capacitors on printed-circuit boards (PCBs). They predominantly use BJTs, and their design differs in significant ways from the design of integrated-circuit (IC) amplifiers. The latter predominantly use MOSFETs, and their study begins in Chapter 8. However, the fundamental principles and concepts introduced in this chapter apply equally well to both discrete and integrated amplifiers.

7.1 Basic Principles

7.1.1 The Basis for Amplifier Operation

The basis for using the transistor (a MOSFET or a BJT) in amplifier design is that when we operate the device in the active region, we create a voltage-controlled current source. Specifically, when a MOSFET is operated in the saturation or pinch-off region, also referred to in this chapter as the active region, the voltage between gate and source, v_{GS}, controls the drain current i_D according to the square-law relationship. For an NMOS transistor, this is expressed as

$$i_D = \frac{1}{2}k_n(v_{GS} - V_{tn})^2 \tag{7.1}$$

In this first-order model of MOSFET operation, the drain current i_D does *not* depend on the drain voltage v_{DS} because the channel is pinched off at the drain end, thus "isolating" the drain. To focus on essentials, we will neglect the Early effect for now.

Similarly, when a BJT is operated in the active region, the base-emitter voltage v_{BE} controls the collector current i_C according to the exponential relationship which, for an *npn* transistor, is expressed as

$$i_C = I_S e^{v_{BE}/V_T} \tag{7.2}$$

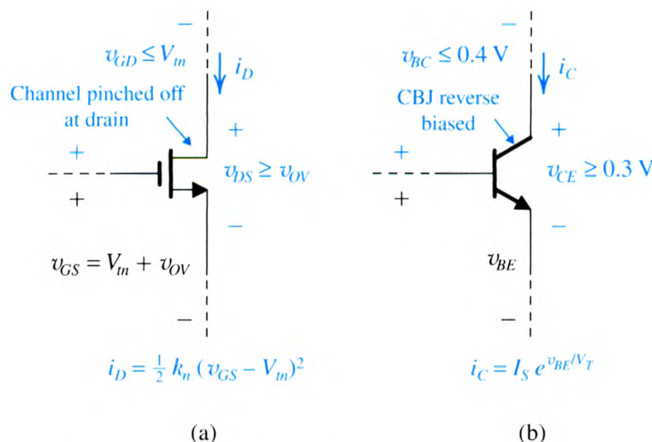


Figure 7.1 Operating (**a**) an NMOS transistor and (**b**) an *npn* transistor in the active mode. Note that $v_{GS} = V_{tn} + v_{OV}$ and $v_{DS} \geq v_{OV}$; thus $v_{GD} \leq V_{tn}$, which ensures channel pinch-off at the drain end. Similarly, $v_{BE} \simeq$ 0.7 V, and $v_{CE} \geq 0.3$ V results in $v_{BC} \leq 0.4$ V, which is sufficient to keep the CBJ from conducting.

Here, this first-order model of BJT operation indicates that the collector current i_C does *not* depend on the collector voltage v_{CE} because the collector–base junction is reverse biased, thus "isolating" the collector.

Figure 7.1 shows an NMOS transistor and an *npn* transistor operating in the active mode. Notice that for the NMOS transistor, the pinch-off condition is ensured by keeping $v_{DS} \geq v_{OV}$. Since the overdrive voltage $v_{OV} = v_{GS} - V_{tn}$, this condition implies that $v_{GD} \leq V_{tn}$, which indeed ensures channel pinch-off at the drain end.

Similarly, for the *npn* transistor in Fig. 7.1(b), the CBJ reverse-bias condition is ensured by keeping $v_{CE} \geq 0.3$ V. Since v_{BE} is usually in the vicinity of 0.7 V, v_{BC} is kept smaller than 0.4 V, which is enough to prevent this relatively large-area junction from conducting.

Although we used NMOS and *npn* transistors to illustrate the conditions for active-mode operation, similar conditions apply for PMOS and *pnp* transistors, as we studied in Chapters 5 and 6, respectively.

Finally, we note that the control relationships in Eqs. (7.1) and (7.2) are nonlinear. Nevertheless, we shall shortly devise a technique for obtaining almost-linear amplification from these fundamentally nonlinear devices.

7.1.2 Obtaining a Voltage Amplifier

From the above we see that the transistor is basically a transconductance amplifier: that is, an amplifier whose input signal is a voltage and whose output signal is a current. More commonly, however, we are interested in voltage amplifiers. A simple way to convert a transconductance amplifier to a voltage amplifier is to pass the output current through a resistor and take the voltage across the resistor as the output. Doing this for a MOSFET gives us the simple amplifier circuit shown in Fig. 7.2(a). Here v_{GS} is the input voltage, R_D (known as a **load resistance**) converts the drain current i_D to a voltage ($i_D R_D$), and V_{DD} is the supply voltage that powers up the amplifier and, together with R_D, establishes operation in the active region, as we will show shortly.

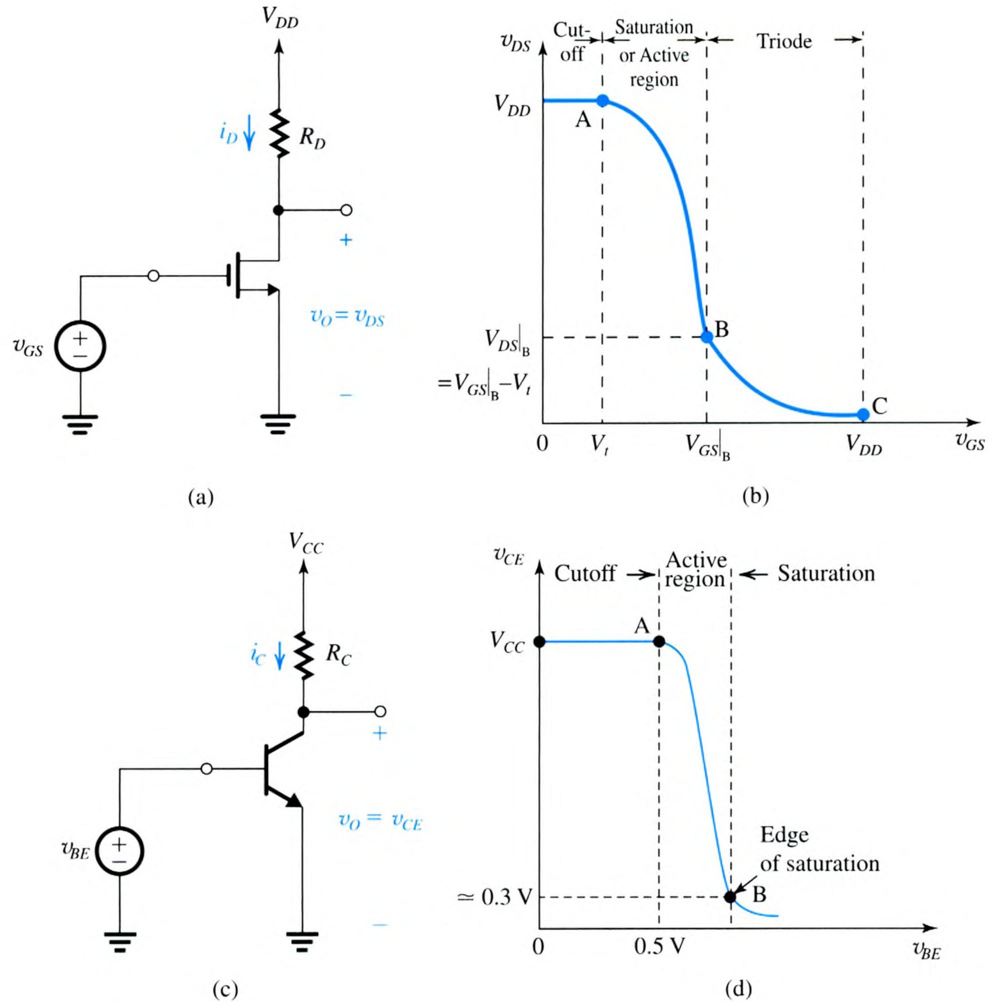


Figure 7.2 (**a**) An NMOS amplifier and (**b**) its voltage transfer characteristic (VTC); and (**c**) an *npn* amplifier and (**d**) its VTC.

In the amplifier circuit of Fig. 7.2(a) the output voltage is taken between the drain and ground, rather than simply across R_D. This is because we need to maintain a common ground reference between the input and the output. The output voltage v_{DS} is given by

$$v_{DS} = V_{DD} - i_D R_D \tag{7.3}$$

It is an inverted version (note the minus sign) of $i_D R_D$ that is shifted by the constant value of the supply voltage V_{DD}.

An exactly similar arrangement applies for the BJT amplifier, as illustrated in Fig. 7.2(c). Here the output voltage v_{CE} is given by

$$v_{CE} = V_{CC} - i_C R_C \tag{7.4}$$

7.1.3 The Voltage-Transfer Characteristic (VTC)

A useful tool that provides insight into the operation of an amplifier circuit is its voltage-transfer characteristic (VTC). This is simply a plot (or a clearly labeled sketch) of the output voltage versus the input voltage. For the MOS amplifier in Fig. 7.2(a), this is the plot of v_{DS} versus v_{GS} shown in Fig. 7.2(b).

Observe that for $v_{GS} < V_t$, the transistor is cut off, $i_D = 0$ and, from Eq. (7.3), $v_{DS} = V_{DD}$. As v_{GS} exceeds V_t, the transistor turns on and v_{DS} decreases. However, since initially v_{DS} is still high, the MOSFET operates in saturation or the active region. This continues as v_{GS} is increased until the value of v_{GS} is reached that results in v_{DS} becoming lower than v_{GS} by V_t volts [point B on the VTC in Fig. 7.2(b)]. For v_{GS} greater than that at point B, the transistor operates in the triode region and v_{DS} decreases more slowly.

The VTC in Fig. 7.2(b) indicates that the segment of greatest slope (hence potentially the largest amplifier gain) is that labeled AB, which corresponds to operation in the active region. To use a MOSFET as an amplifier, we confine its operating point to the segment AB at all times. An expression for the segment AB can be obtained by substituting for i_D in Eq. (7.3) by its active-region value from Eq. (7.1), thus

$$v_{DS} = V_{DD} - \frac{1}{2}k_n R_D (v_{GS} - V_t)^2 \tag{7.5}$$

This is obviously a nonlinear relationship. Nevertheless, linear (or almost-linear) amplification can be obtained by using the technique of biasing the MOSFET. Before considering biasing, however, it is useful to determine the coordinates of point B, which is at the boundary between the saturation and the triode regions of operation. These can be obtained by substituting in Eq. (7.5), $v_{GS} = V_{GS}|_B$ and $v_{DS} = V_{DS}|_B = V_{GS}|_B - V_t$. The result is

$$V_{GS}|_B = V_t + \frac{\sqrt{2k_n R_D V_{DD} + 1} - 1}{k_n R_D} \tag{7.6}$$

Point B can be alternatively characterized by the overdrive voltage

$$V_{OV}|_B \equiv V_{GS}|_B - V_t = \frac{\sqrt{2k_n R_D V_{DD} + 1} - 1}{k_n R_D} \tag{7.7}$$

and

$$V_{DS}|_B = V_{OV}|_B \tag{7.8}$$

EXERCISE

7.1 Consider the amplifier of Fig. 7.2(a) with $V_{DD} = 1.8$ V, $R_D = 17.5$ kΩ, and with a MOSFET specified to have $V_t = 0.4$ V, $k_n = 4$ mA/V^2, and $\lambda = 0$. Determine the coordinates of the end points of the active-region segment of the VTC. Also, determine $V_{DS}|_C$ assuming $V_{GS}|_C = V_{DD}$.

Ans. A: 0.4 V, 1.8 V; B: 0.613 V, 0.213 V; $V_{DS}|_C = 18$ mV

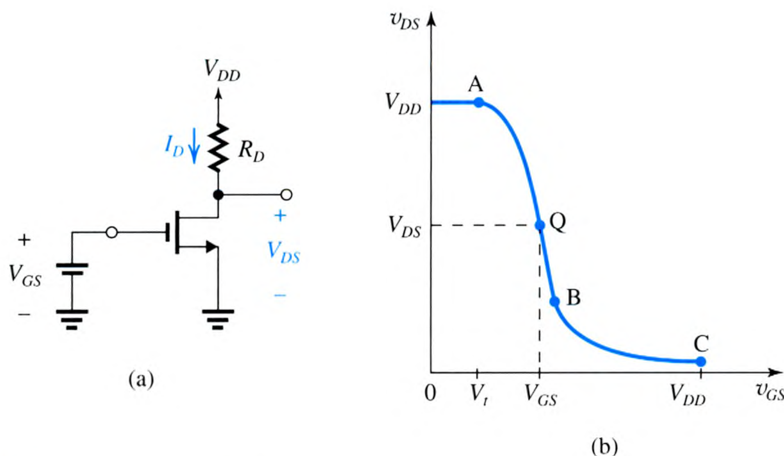


Figure 7.3 Biasing the MOSFET amplifier at a point Q located on the segment AB of the VTC.

An exactly similar development applies to the BJT case, as illustrated in Fig. 7.2(c) and (d). In this case, over the active-region or amplifier segment AB, the output voltage v_{CE} is related to the input voltage v_{BE} by

$$v_{CE} = V_{CC} - R_C I_S e^{v_{BE}/V_T} \tag{7.9}$$

Here also, the input–output relationship is nonlinear. Nevertheless, linear (or almost-linear) amplification can be obtained by using the biasing technique discussed next.

7.1.4 Obtaining Linear Amplification by Biasing the Transistor

Biasing enables us to obtain almost-linear amplification from the MOSFET and the BJT. The technique is illustrated for the MOSFET case in Fig. 7.3(a). We select a dc voltage V_{GS} to obtain operation at a point Q on the segment AB of the VTC. We will discuss how to select an appropriate location for the bias point Q shortly. For the time being, observe that the coordinates of Q are the dc voltages V_{GS} and V_{DS}, which are related by

$$V_{DS} = V_{DD} - \frac{1}{2}k_n R_D (V_{GS} - V_t)^2 \tag{7.10}$$

Point Q is known as the **bias point** or the **dc operating point**. Since at Q no signal component is present, it is also known as the **quiescent point** (hence the symbol Q).

Next, the signal to be amplified, v_{gs}, a function of time t, is superimposed on the bias voltage V_{GS}, as shown in Fig. 7.4(a). Thus the total instantaneous value of v_{GS} becomes

$$v_{GS}(t) = V_{GS} + v_{gs}(t)$$

The resulting $v_{DS}(t)$ can be obtained by substituting for $v_{GS}(t)$ into Eq. (7.5). Graphically, we can use the VTC to obtain $v_{DS}(t)$ point by point, as illustrated in Fig. 7.4(b). Here we show v_{gs} as a triangular wave of "small" amplitude. Specifically, the amplitude of v_{gs} is small enough to restrict the excursion of the instantaneous operating point to a short, almost-linear segment of the VTC around the bias point Q. The shorter the segment, the greater the linearity achieved, and the closer to an ideal triangular wave the signal component at the output, v_{ds}, will be. This is the essence of obtaining linear amplification from the nonlinear MOSFET.

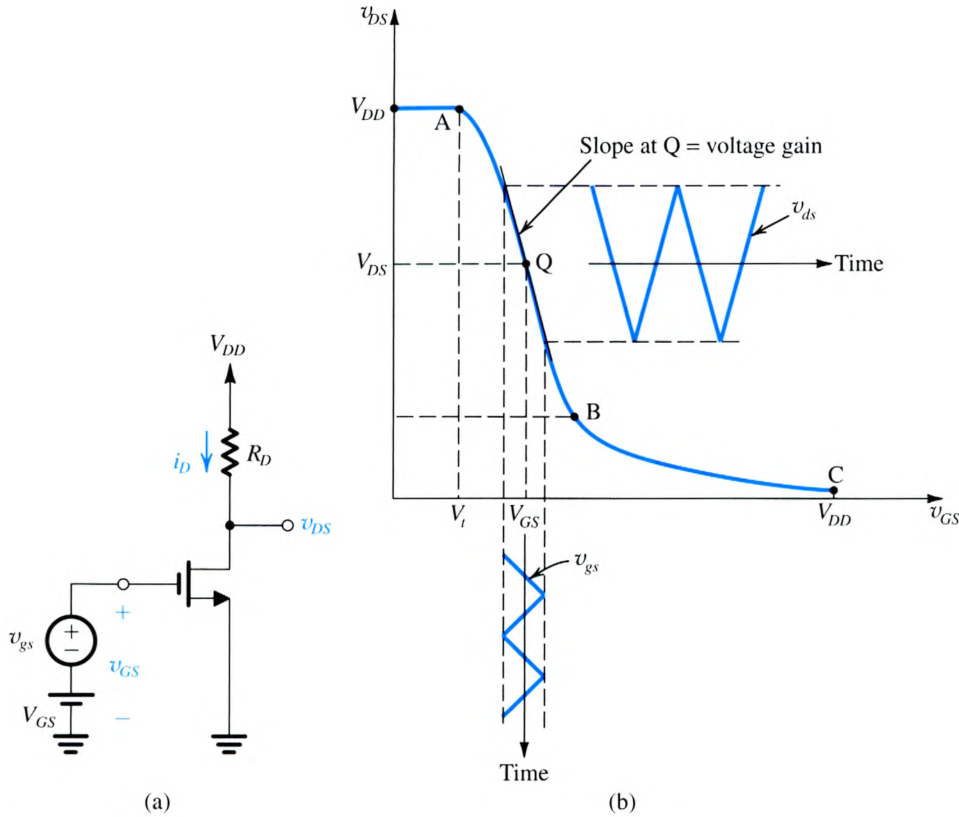


Figure 7.4 **(a)** The MOSFET amplifier with a small time-varying signal $v_{gs}(t)$ superimposed on the dc bias voltage V_{GS}. **(b)** The MOSFET operates on a short, almost-linear segment of the VTC around the bias point Q and provides an output voltage $v_{ds} = A_v v_{gs}$.

Before leaving Fig. 7.4(b) let's look at what happens when we increase the amplitude of the signal v_{gs}. As the instantaneous operating point is no longer confined to the almost-linear segment of the VTC, the output signal v_{ds} deviates from its ideal triangular shape; that is, it exhibits nonlinear distortion. Worse yet, if the input signal amplitude becomes sufficiently large, the instantaneous operating point may leave the segment AB altogether. If this happens at the negative peaks of v_{gs}, the transistor will cut off for a portion of the cycle and the positive peaks of v_{ds} will be "clipped off." If it occurs at the positive peaks of v_{gs}, the transistor will enter the triode region for a portion of the cycle, and the negative peaks of v_{ds} will become flattened. It follows that the selection of the location of the bias point Q can have a profound effect on the maximum allowable amplitude of v_{ds}, referred to as the *allowable signal swing at the output*. We will have more to say later on this important point.

An exactly parallel development can be applied to the BJT amplifier. In fact, all we need to do is replace the NMOS transistor in Figs. 7.3 and 7.4 with an *npn* transistor and change the voltage and current symbols to their BJT counterparts. The resulting bias point Q will be characterized by dc voltages V_{BE} and V_{CE}, which are related by

$$V_{CE} = V_{CC} - R_C I_S e^{V_{BE}/V_T} \tag{7.11}$$

and a dc current I_C,

$$I_C = I_S e^{V_{BE}/V_T} \tag{7.12}$$

Also, superimposing a small-signal v_{be} on the dc bias voltage V_{BE} results in

$$v_{BE}(t) = V_{BE} + v_{be}(t)$$

which can be substituted into Eq. (7.9) to obtain the total instantaneous value of the output voltage $v_{CE}(t)$. Here again, we get almost-linear operation by keeping v_{be} small enough to restrict the excursion of the instantaneous operating point to a short, almost-linear segment of the VTC around the bias point Q. Similar comments also apply to the maximum allowable signal swing at the output.

7.1.5 The Small-Signal Voltage Gain

The MOSFET Case Consider the MOSFET amplifier in Fig. 7.4(a). If the input signal v_{gs} is kept small, the corresponding signal at the output v_{ds} will be nearly proportional to v_{gs} with the constant of proportionality being the slope of the almost-linear segment of the VTC around Q. This is the voltage gain of the amplifier, and we can determine its value by finding the slope of the tangent to the VTC at the bias point Q,

$$A_v = \left.\frac{dv_{DS}}{dv_{GS}}\right|_{v_{GS}=V_{GS}} \tag{7.13}$$

Using Eq. (7.5) we obtain

$$A_v = -k_n(V_{GS} - V_t)R_D \tag{7.14}$$

which we can express in terms of the overdrive voltage at the bias point, V_{OV}, as

$$A_v = -k_n V_{OV} R_D \tag{7.15}$$

We make the following observations on this expression for the voltage gain.

1. The gain is negative, which tells us that the amplifier is inverting; that is, there is a 180° phase shift between the input and the output. This inversion is obvious in Fig. 7.4(b) and should have been anticipated from Eq. (7.5).
2. The gain is proportional to the load resistance R_D, to the transistor transconductance parameter k_n, and to the overdrive voltage V_{OV}. This all makes intuitive sense.

Another simple and insightful expression for the voltage gain A_v can be derived by recalling that the dc current in the drain at the bias point is related to V_{OV} by

$$I_D = \frac{1}{2}k_n V_{OV}^2$$

This equation can be combined with Eq. (7.15) to obtain

$$A_v = -\frac{I_D R_D}{V_{OV}/2} \tag{7.16}$$

That is, the gain is simply the ratio of the dc voltage drop across the load resistance R_D to $V_{OV}/2$. We can express it in the alternative form

$$A_v = -\frac{V_{DD} - V_{DS}}{V_{OV}/2} \tag{7.17}$$

Since the maximum slope of the VTC in Fig. 7.4(b) occurs at point B, the maximum gain magnitude $|A_{v\max}|$ is obtained by biasing the transistor at point B,

$$|A_{v\max}| = \frac{V_{DD} - V_{DS}\big|_B}{V_{OV}\big|_B/2}$$

and since $V_{DS}\big|_B = V_{OV}\big|_B$,

$$|A_{v\max}| = \frac{V_{DD} - V_{OV}\big|_B}{V_{OV}\big|_B/2} \tag{7.18}$$

where $V_{OV}\big|_B$ is given by Eq. (7.7). Of course, this result is only of theoretical importance since biasing at B would leave no room for negative signal swing at the output. Nevertheless, the result in Eq. (7.18) is valuable as it tells us the maximum voltage gain achievable from this basic amplifier circuit. As an example, for a discrete-circuit amplifier operated with $V_{DD} = 5$ V and $V_{OV}\big|_B = 0.5$ V, the maximum achievable gain is 18 V/V. An integrated-circuit amplifier using a modern submicron MOSFET operated with $V_{DD} = 1.3$ V and with $V_{OV}\big|_B = 0.2$ V realizes a maximum gain of 11 V/V.

Finally, note that to maximize the gain, the bias point Q should be as close to point B as possible, consistent with the required signal swing at the output. We will explore this point further in the end-of-chapter problems.

Example 7.1

Consider the amplifier circuit shown in Fig. 7.4(a). The transistor is specified to have $V_t = 0.4$ V, $k_n' = 0.4$ mA/V^2, $W/L = 10$, and $\lambda = 0$. Also, let $V_{DD} = 1.8$ V, $R_D = 17.5$ kΩ, and $V_{GS} = 0.6$ V.

(a) For $v_{gs} = 0$ (and hence $v_{ds} = 0$), find V_{OV}, I_D, V_{DS}, and A_v.

(b) What is the maximum symmetrical signal swing allowed at the drain? Hence, find the maximum allowable amplitude of a sinusoidal v_{gs}.

Solution

(a) With $V_{GS} = 0.6$ V, $V_{OV} = 0.6 - 0.4 = 0.2$ V. Thus,

$$\begin{aligned} I_D &= \frac{1}{2}k_n'\left(\frac{W}{L}\right)V_{OV}^2 \\ &= \frac{1}{2} \times 0.4 \times 10 \times 0.2^2 = 0.08 \text{ mA} \\ V_{DS} &= V_{DD} - R_D I_D \\ &= 1.8 - 17.5 \times 0.08 = 0.4 \text{ V} \end{aligned}$$

Example 7.1 *continued*

Since V_{DS} is greater than V_{OV}, the transistor is indeed operating in saturation. The voltage gain can be found from Eq. (7.15),

$$\begin{aligned} A_v &= -k_n V_{OV} R_D \\ &= -0.4 \times 10 \times 0.2 \times 17.5 = -14 \text{ V/V} \end{aligned}$$

An identical result can be found using Eq. (7.17).

(b) Since $V_{OV} = 0.2$ V and $V_{DS} = 0.4$ V, we see that the maximum allowable negative signal swing at the drain is 0.2 V. In the positive direction, a swing of +0.2 V would not cause the transistor to cut off (since the resulting v_{DS} would be still lower than V_{DD}) and thus is allowed. Thus the maximum symmetrical signal swing allowable at the drain is ±0.2 V. The corresponding amplitude of v_{gs} can be found from

$$\hat{v}_{gs} = \frac{\hat{v}_{ds}}{|A_v|} = \frac{0.2 \text{ V}}{14} = 14.2 \text{ mV}$$

Since $\hat{v}_{gs} \ll V_{OV}$, the operation will be reasonably linear (more on this in later sections).

Greater insight into the issue of allowable signal swing can be obtained by examining the signal waveforms shown in Fig. 7.5. Note that for the MOSFET to remain in saturation at the negative peak of v_{ds}, we must ensure that

$$v_{DS\min} \geq v_{GS\max} - V_t$$

that is,

$$0.4 - |A_v|\hat{v}_{gs} \geq 0.6 + \hat{v}_{gs} - 0.4$$

which results in

$$\hat{v}_{gs} \leq \frac{0.2}{|A_v| + 1} = 13.3 \text{ mV}$$

This result differs slightly from the one we obtained earlier.

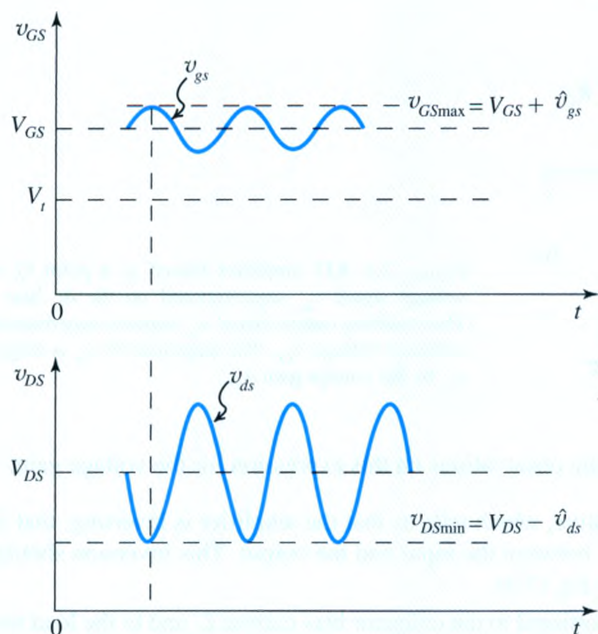


Figure 7.5 Signal waveforms at gate and drain for the amplifier in Example 7.1. Note that to ensure operation in the saturation region at all times, $v_{DS\min} \geq v_{GS\max} - V_t$.

EXERCISE

D7.2 For the amplifier circuit studied in Example 7.1, create two alternative designs, each providing a voltage gain of -10 by (a) changing R_D while keeping V_{OV} constant and (b) changing V_{OV} while keeping R_D constant. For each design, specify V_{GS}, I_D, R_D, and V_{DS}.
Ans. (a) 0.6 V, 0.08 mA, 12.5 kΩ, 0.8 V; (b) 0.54 V, 0.04 mA, 17.5 kΩ, 1.1 V

The BJT Case We can use a similar development to obtain the small-signal voltage gain of the BJT amplifier shown in Fig. 7.6,

$$A_v = \left.\frac{dv_{CE}}{dv_{BE}}\right|_{v_{BE}=V_{BE}} \tag{7.19}$$

Using Eq. (7.9) together with Eq. (7.12), we obtain

$$A_v = -\left(\frac{I_C}{V_T}\right)R_C \tag{7.20}$$

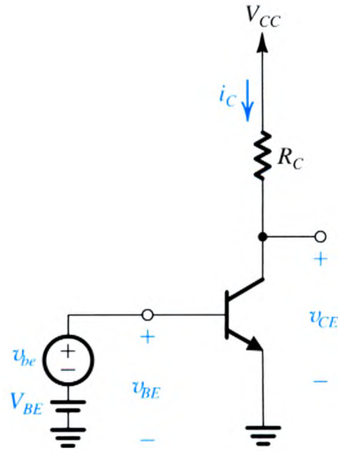


Figure 7.6 BJT amplifier biased at a point Q, with a small voltage signal v_{be} superimposed on the dc bias voltage V_{BE}. The resulting output signal v_{ce} appears superimposed on the dc collector voltage V_{CE}. The amplitude of v_{ce} is larger than that of v_{be} by the voltage gain A_v.

We make the following observations on this expression for the voltage gain:

1. The gain is negative, which tells us that the amplifier is inverting; that is, there is a 180° phase shift between the input and the output. This inversion should have been anticipated from Eq. (7.9).
2. The gain is proportional to the collector bias current I_C and to the load resistance R_C.

Additional insight into the voltage gain A_v can be obtained by expressing Eq. (7.20) as

$$A_v = -\frac{I_C R_C}{V_T} \tag{7.21}$$

That is, the gain is the ratio of the dc voltage drop across the load resistance R_C to the physical constant V_T (recall that the thermal voltage $V_T \simeq 25$ mV at room temperature). This relationship is similar in form to that for the MOSFET (Eq. 7.16) except that here the denominator is a physical constant (V_T) rather than a design parameter ($V_{OV}/2$). Usually, $V_{OV}/2$ is larger than V_T, thus we can obtain higher voltage gain from the BJT amplifier than from the MOSFET amplifier. This should not be surprising, as the exponential i_C–v_{BE} relationship is much steeper than the square-law relationship i_D–v_{GS}.

The gain A_v in Eq. (7.21) can be expressed alternately as

$$A_v = -\frac{V_{CC} - V_{CE}}{V_T} \tag{7.22}$$

from which we see that maximum gain is achieved when V_{CE} is at its minimum value of about 0.3 V,

$$|A_{v\max}| = \frac{V_{CC} - 0.3}{V_T} \tag{7.23}$$

Here again, this is only a theoretical maximum, since biasing the BJT at the edge of saturation leaves no room for negative signal swing at the output. Equation (7.23) nevertheless tells us the maximum voltage gain achievable from the basic BJT amplifier. As an example, for $V_{CC} = 5$ V, the maximum gain is 188 V/V, considerably larger than in the MOSFET case. For modern low-voltage technologies, a V_{CC} of 1.3 V provides a gain of 40 V/V, again much larger than the MOSFET case. The reader should not, however,

jump to the conclusion that the BJT is preferred to the MOSFET in the design of modern integrated-circuit amplifiers; in fact, the opposite is true, as we shall see in Chapter 8 and beyond.

Finally, Eq. (7.22) shows us that to maximize $|A_v|$ the transistor should be biased at the lowest possible V_{CE} consistent with the desired value of negative signal swing at the output.

Example 7.2

Consider an amplifier circuit using a BJT having $I_S = 10^{-15}$ A, a collector resistance $R_C = 6.8$ kΩ, and a power supply $V_{CC} = 10$ V.

(a) Determine the value of the bias voltage V_{BE} required to operate the transistor at $V_{CE} = 3.2$ V. What is the corresponding value of I_C?

(b) Find the voltage gain A_v at this bias point. If an input sine-wave signal of 5-mV peak amplitude is superimposed on V_{BE}, find the amplitude of the output sine-wave signal (assume linear operation).

(c) Find the positive increment in v_{BE} (above V_{BE}) that drives the transistor to the edge of saturation, where $v_{CE} = 0.3$ V.

(d) Find the negative increment in v_{BE} that drives the transistor to within 1% of cutoff (i.e., to $v_{CE} = 0.99V_{CC}$).

Solution

(a)

$$I_C = \frac{V_{CC} - V_{CE}}{R_C}$$

$$= \frac{10 - 3.2}{6.8} = 1 \text{ mA}$$

The value of V_{BE} can be determined from

$$1 \times 10^{-3} = 10^{-15} e^{V_{BE}/V_T}$$

which results in

$$V_{BE} = 690.8 \text{ mV}$$

(b)

$$A_v = -\frac{V_{CC} - V_{CE}}{V_T}$$

$$= -\frac{10 - 3.2}{0.025} = -272 \text{ V/V}$$

$$\hat{v}_{ce} = 272 \times 0.005 = 1.36 \text{ V}$$

(c) For $v_{CE} = 0.3$ V,

$$i_C = \frac{10 - 0.3}{6.8} = 1.426 \text{ mA}$$

Example 7.2 *continued*

To increase i_C from 1 mA to 1.426 mA, v_{BE} must be increased by

$$\Delta v_{BE} = V_T \ln\left(\frac{1.426}{1}\right) = 9\text{ mV}$$

(d) For $v_{CE} = 0.99V_{CC} = 9.9$ V,

$$i_C = \frac{10 - 9.9}{6.8} = 0.0147\text{ mA}$$

To decrease i_C from 1 mA to 0.0147 mA, v_{BE} must change by

$$\Delta v_{BE} = V_T \ln\left(\frac{0.0147}{1}\right) = -105.5\text{ mV}$$

EXERCISE

7.3 For the situation described in Example 7.2, while keeping I_C unchanged at 1 mA, find the value of R_C that will result in a voltage gain of -320 V/V. What is the largest negative signal swing allowed at the output (assume that v_{CE} is not to decrease below 0.3 V)? What (approximately) is the corresponding input signal amplitude? (Assume linear operation.)
Ans. 8 kΩ; 1.7 V; 5.3 mV

7.1.6 Determining the VTC by Graphical Analysis

Figure 7.7 shows a graphical method for determining the VTC of the amplifier of Fig. 7.4(a). Although graphical analysis of transistor circuits is rarely employed in practice, it is useful to us at this stage for gaining greater insight into circuit operation, especially in figuring out where to locate the bias point Q.

The graphical analysis is based on the observation that for each value of v_{GS}, the circuit will be operating at the point of intersection of the i_D-v_{DS} graph corresponding to the particular value of v_{GS} and the straight line representing Eq. (7.3), which can be rewritten in the form

$$i_D = \frac{V_{DD}}{R_D} - \frac{1}{R_D} v_{DS} \tag{7.24}$$

The straight line representing this relationship is superimposed on the i_D-v_{DS} characteristics in Fig. 7.7. It intersects the horizontal axis at $v_{DS} = V_{DD}$ and has a slope of $-1/R_D$. Since this

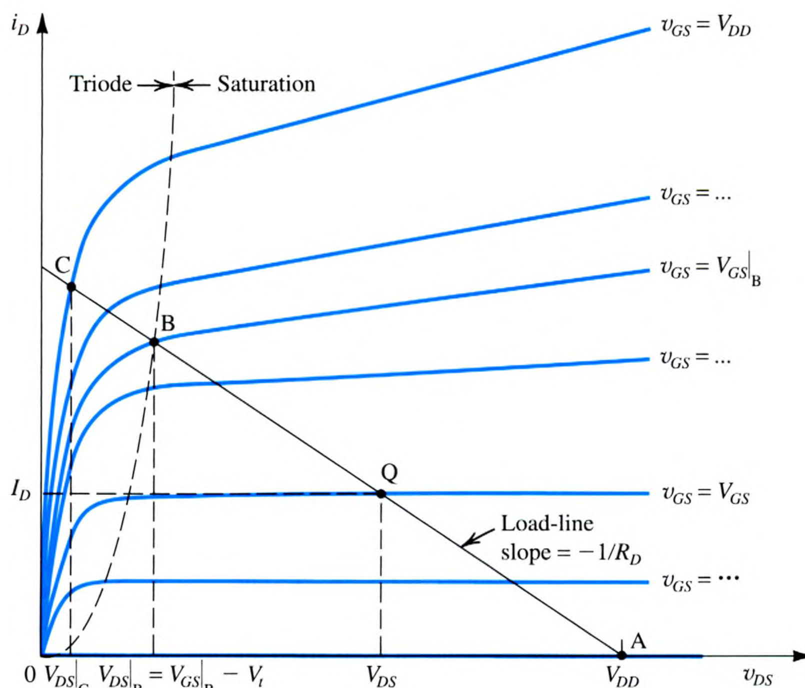


Figure 7.7 Graphical construction to determine the voltage-transfer characteristic of the amplifier in Fig. 7.4(a).

straight line represents in effect the load resistance R_D, it is called the **load line**. The VTC is then determined point by point. Note that we have labeled four important points: point A at which $v_{GS} = V_t$, point Q at which the MOSFET can be biased for amplifier operation ($v_{GS} = V_{GS}$ and $v_{DS} = V_{DS}$), point B at which the MOSFET leaves saturation and enters the triode region, and point C, which is deep into the triode region and for which $v_{GS} = V_{DD}$. If the MOSFET is to be used as a switch, then operating points A and C are applicable: At A the transistor is off (open switch), and at C the transistor operates as a low-valued resistance r_{DS} and has a small voltage drop (closed switch). The incremental resistance at point C is also known as the **closure resistance**. The operation of the MOSFET as a switch is illustrated in Fig. 7.8. A detailed study of the application of the MOSFET as a switch is undertaken in Chapter 16, dealing with CMOS digital logic circuits.

We can use the graphical analysis method above to determine the VTC of the BJT amplifier in Fig. 7.2(c). Here point A, Fig. 7.2(d), corresponds to the BJT just turning on ($v_{BE} \simeq 0.5$ V), and point B corresponds to the BJT leaving the active region and entering the saturation region. If the BJT is to be operated as a switch, the two modes of operation are cutoff (open switch) and saturation (closed switch). As discussed in Section 6.2, in saturation, the BJT has a small closure resistance R_{CEsat} as well as an offset voltage. More seriously, switching the BJT out of its saturation region can require a relatively long delay time to ensure the removal of the charge stored in the BJT base region. This phenomenon has made the BJT much less attractive than the MOSFET in digital logic applications.[1]

[1]The only exception is a nonsaturating form of BJT logic circuits known as emitter-coupled logic (ECL).

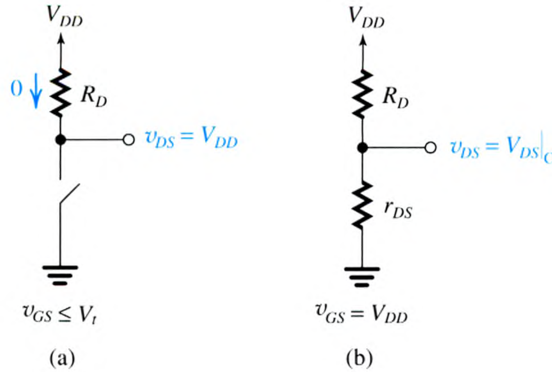


Figure 7.8 Operation of the MOSFET in Fig. 7.4(a) as a switch: **(a)** open, corresponding to point A in Fig. 7.7; **(b)** closed, corresponding to point C in Fig. 7.7. The closure resistance is approximately equal to r_{DS} because V_{DS} is usually very small.

7.1.7 Deciding on a Location for the Bias Point Q

For the MOSFET amplifier, the bias point Q is determined by the values of V_{GS} and of the load resistance R_D. Two important considerations in deciding on the location of Q are the required gain and the desired signal swing at the output. To illustrate this, consider the VTC shown in Fig. 7.4(b). Here the value of R_D is fixed and the only variable remaining is the value of V_{GS}. Since the slope increases as we move closer to point B, we obtain higher gain by locating Q as close to B as possible. However, the closer Q is to the boundary point B, the smaller the allowable magnitude of negative signal swing. Thus, as often happens in engineering design, we need to make a trade-off. The answer here is relatively simple: For a given R_D, locate Q as close to the triode region (point B) as possible to obtain high gain but sufficiently distant to allow for the required negative signal swing. However, when we take the Early effect into account, the determination of an optimal bias point Q can become more complicated.

In deciding on a value for R_D, it helps to refer to the $i_D - v_{DS}$ plane. Figure 7.9 shows two load lines giving two extreme bias points: Point Q_1 is too close to V_{DD}, putting a severe constraint on the positive signal swing of v_{ds}. Exceeding the allowable positive maximum causes the positive peaks of the signal to be clipped off, since the MOSFET will turn off for the part of each cycle near the positive peak. We say that the circuit does not have sufficient "headroom." Similarly, point Q_2 is too close to the boundary of the triode region, thus severely limiting the allowable negative signal swing of v_{ds}. Exceeding this limit would cause the transistor to enter the triode region for part of each cycle near the negative peaks, resulting in a distorted output signal. In this situation, we say that the circuit does not have sufficient "legroom." We will have more to say on bias design in Section 7.4. Finally, note that exactly the same considerations apply to the BJT amplifier.

7.2 Small-Signal Operation and Models

In Section 7.1 we learned that we can obtain linear amplification by biasing the transistor to operate in the active region and by keeping the input signal small. In this section, we explore the small-signal operation in greater detail.

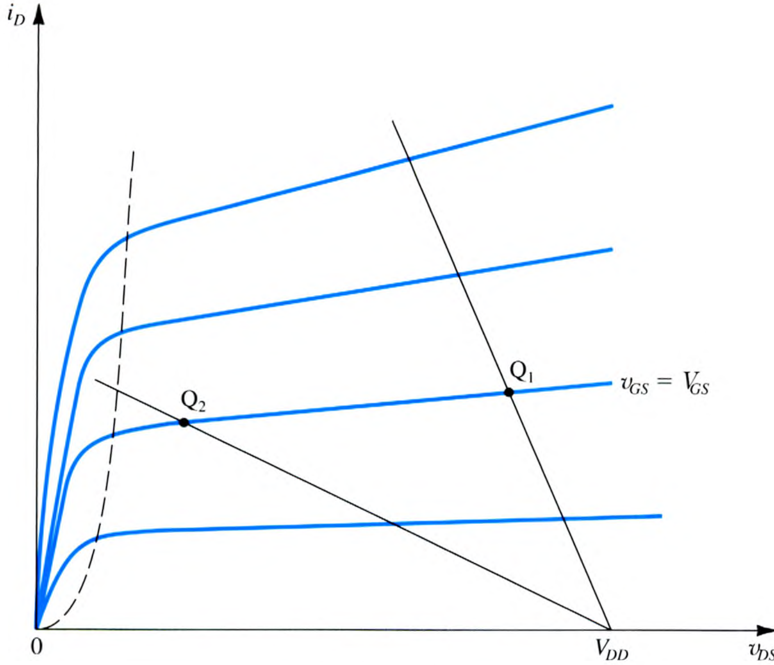


Figure 7.9 Two load lines and corresponding bias points. Bias point Q_1 does not leave sufficient room for positive signal swing at the drain (too close to V_{DD}). Bias point Q_2 is too close to the boundary of the triode region and might not allow for sufficient negative signal swing.

7.2.1 The MOSFET Case

Consider the conceptual amplifier circuit shown in Fig. 7.10. Here the MOS transistor is biased by applying a dc voltage[2] V_{GS}, and the input signal to be amplified, v_{gs}, is superimposed on the dc bias voltage V_{GS}. The output voltage is taken at the drain.

The DC Bias Point We can find the dc bias current I_D by setting the signal v_{gs} to zero; thus,

$$I_D = \frac{1}{2}k_n(V_{GS} - V_t)^2 = \frac{1}{2}k_n V_{OV}^2 \tag{7.25}$$

where we have neglected channel-length modulation (i.e., we have assumed $\lambda = 0$). Here $V_{OV} = V_{GS} - V_t$ is the overdrive voltage at which the MOSFET is biased to operate. The dc voltage at the drain, V_{DS}, will be

$$V_{DS} = V_{DD} - R_D I_D \tag{7.26}$$

To ensure saturation-region operation, we must have

$$V_{DS} > V_{OV}$$

[2]Practical biasing arrangements will be studied in Section 7.4.

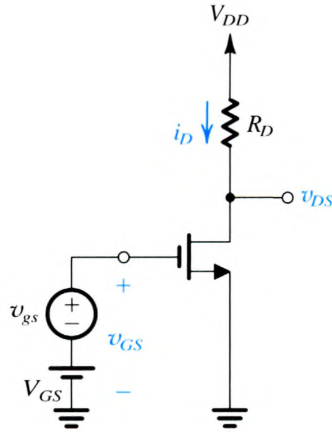


Figure 7.10 Conceptual circuit utilized to study the operation of the MOSFET as a small-signal amplifier.

Furthermore, since the total voltage at the drain will have a signal component superimposed on V_{DS}, V_{DS} has to be sufficiently greater than V_{OV} to allow for the required negative signal swing.

The Signal Current in the Drain Terminal Next, let us apply the input signal v_{gs}. The total instantaneous gate-to-source voltage will be

$$v_{GS} = V_{GS} + v_{gs} \tag{7.27}$$

resulting in a total instantaneous drain current i_D,

$$\begin{aligned} i_D &= \frac{1}{2}k_n\left(V_{GS} + v_{gs} - V_t\right)^2 \\ &= \frac{1}{2}k_n(V_{GS} - V_t)^2 + k_n(V_{GS} - V_t)v_{gs} + \frac{1}{2}k_n v_{gs}^2 \end{aligned} \tag{7.28}$$

The first term on the right-hand side of Eq. (7.28) is the dc bias current I_D (Eq. 7.25). The second term represents a current component that is directly proportional to the input signal v_{gs}. The third term is a current component that is proportional to the square of the input signal. This last component is undesirable because it represents *nonlinear distortion*. To reduce the nonlinear distortion introduced by the MOSFET, the input signal should be kept small so that

$$\frac{1}{2}k_n v_{gs}^2 \ll k_n(V_{GS} - V_t)v_{gs}$$

resulting in

$$v_{gs} \ll 2(V_{GS} - V_t) \tag{7.29}$$

or, equivalently,

$$v_{gs} \ll 2V_{OV} \tag{7.30}$$

If this **small-signal condition** is satisfied, we may neglect the last term in Eq. (7.28) and express i_D as

$$i_D \simeq I_D + i_d \tag{7.31}$$

where

$$i_d = k_n(V_{GS} - V_t)v_{gs}$$

The parameter that relates i_d and v_{gs} is the MOSFET **transconductance** g_m,

$$g_m \equiv \frac{i_d}{v_{gs}} = k_n(V_{GS} - V_t) \tag{7.32}$$

or in terms of the overdrive voltage V_{OV},

$$g_m = k_n V_{OV} \tag{7.33}$$

Figure 7.11 presents a graphical interpretation of the small-signal operation of the MOSFET amplifier. Note that g_m is equal to the slope of the i_D–v_{GS} characteristic at the bias point,

$$g_m \equiv \left.\frac{\partial i_D}{\partial v_{GS}}\right|_{v_{GS}=V_{GS}} \tag{7.34}$$

This is the formal definition of g_m, which can be shown to yield the expressions given in Eqs. (7.32) and (7.33).

The Voltage Gain Returning to the circuit in Fig. 7.10, we can express the total instantaneous drain voltage v_{DS} as follows:

$$v_{DS} = V_{DD} - R_D i_D$$

Under the small-signal condition, we have

$$v_{DS} = V_{DD} - R_D(I_D + i_d)$$

which we can rewrite as

$$v_{DS} = V_{DS} - R_D i_d$$

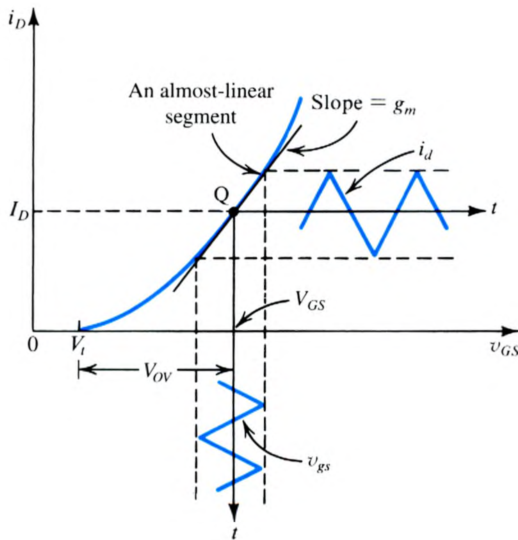


Figure 7.11 Small-signal operation of the MOSFET amplifier.

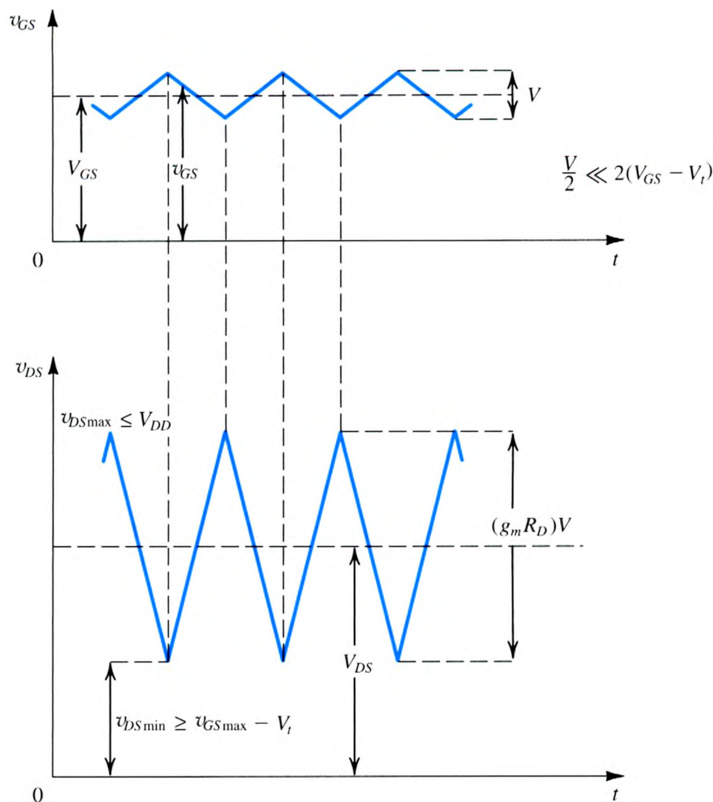


Figure 7.12 Total instantaneous voltages v_{GS} and v_{DS} for the circuit in Fig. 7.10.

Thus the signal component of the drain voltage is

$$v_{ds} = -i_d R_D = -g_m v_{gs} R_D \tag{7.35}$$

which indicates that the voltage gain is given by

$$A_v \equiv \frac{v_{ds}}{v_{gs}} = -g_m R_D \tag{7.36}$$

The minus sign in Eq. (7.36) indicates that the output signal v_{ds} is 180° out of phase with the input signal v_{gs}. This is illustrated in Fig. 7.12, which shows v_{GS} and v_{DS}. The input signal is assumed to have a triangular waveform with an amplitude much smaller than $2(V_{GS} - V_t)$, the small-signal condition in Eq. (7.29), to ensure linear operation. For operation in the saturation (active) region at all times, the minimum value of v_{DS} should not fall below the corresponding value of v_{GS} by more than V_t. Also, the maximum value of v_{DS} should be smaller than V_{DD}; otherwise the FET will enter the cutoff region and the peaks of the output signal waveform will be clipped off.

Finally, notice that by substituting for g_m from Eq. (7.33) the voltage-gain expression in Eq. (7.36) becomes identical to that derived in Section 7.1—namely, Eq. (7.15).

Separating the DC Analysis and the Signal Analysis From the preceding analysis, we see that under the small-signal approximation, signal quantities are superimposed on dc

quantities. For instance, the total drain current i_D equals the dc current I_D plus the signal current i_d, the total drain voltage $v_{DS} = V_{DS} + v_{ds}$, and so on. It follows that we can greatly simplify the analysis and design by separating dc or bias calculations from small-signal calculations. That is, once we have established a stable dc operating point and have calculated all dc quantities, we can perform signal analysis ignoring dc quantities.

Small-Signal Equivalent-Circuit Models From a signal point of view, the FET behaves as a voltage-controlled current source. It accepts a signal v_{gs} between its gate and source and provides a current $g_m v_{gs}$ at the drain terminal. The input resistance of this controlled source is very high—ideally, infinite. The output resistance—that is, the resistance looking into the drain—also is high, and so far we have assumed it to be infinite. Putting all of this together, we arrive at the circuit in Fig. 7.13(a), which represents the small-signal operation of the MOSFET and is thus a **small-signal model** or a **small-signal equivalent circuit**.

In analyzing a MOSFET amplifier circuit, we can replace the transistor with the equivalent-circuit model shown in Fig. 7.13(a). The rest of the circuit remains unchanged except that *ideal constant dc voltage sources are replaced by short circuits*. This is because the voltage across an ideal constant dc voltage source does not change, and thus there will always be a zero voltage signal across a constant dc voltage source. A dual statement applies for constant dc current sources; namely, the signal current of an ideal constant dc current source will always be zero, and thus *an ideal constant dc current source can be replaced by an open circuit* in the small-signal equivalent circuit of the amplifier. We can then use the resulting circuit to perform any required signal analysis, such as calculating voltage gain.

The most serious shortcoming of the small-signal model of Fig. 7.13(a) is that it assumes the drain current in saturation to be independent of the drain voltage. From our study of the MOSFET characteristics in saturation, we know that the drain current does in fact depend on v_{DS} in a linear manner. Such dependence was modeled by a finite resistance r_o between drain and source, whose value was given by Eq. (5.27) in Section 5.2.4, which we repeat here (with the prime on I_D dropped) as

$$r_o = \frac{|V_A|}{I_D} \tag{7.37}$$

where the Early voltage $V_A = 1/\lambda$ is a MOSFET parameter that either is specified or can be measured. Remember that for a given process technology, V_A is proportional to the MOSFET

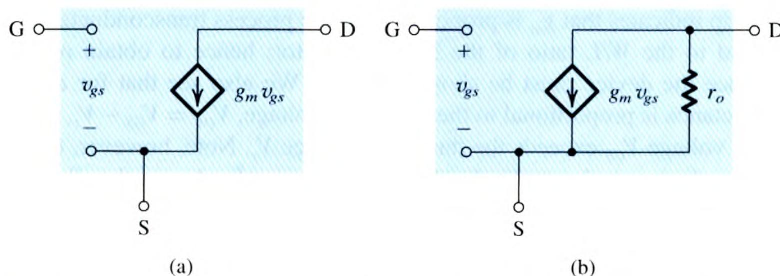


Figure 7.13 Small-signal models for the MOSFET: **(a)** neglecting the dependence of i_D on v_{DS} in the active region (the channel-length modulation effect) and **(b)** including the effect of channel-length modulation, modeled by output resistance $r_o = |V_A|/I_D$. These models apply equally well for both NMOS and PMOS transistors.

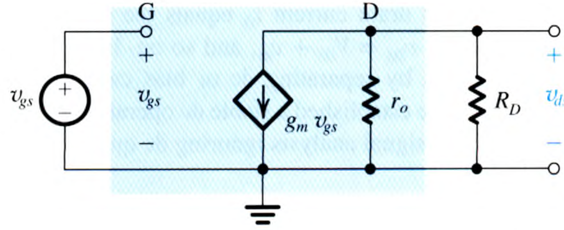


Figure 7.14 Small-signal equivalent circuit model of the amplifier in Fig. 7.10.

channel length. The current I_D is the value of the dc drain current without the channel-length modulation taken into account; that is,

$$I_D = \frac{1}{2}k_n V_{OV}^2 \tag{7.38}$$

Typically, r_o is in the range of 10 kΩ to 1000 kΩ. It follows that we can improve the accuracy of the small-signal model by including r_o in parallel with the controlled source, as shown in Fig. 7.13(b). It is important to note that the small-signal model parameters g_m and r_o depend on the dc bias point of the MOSFET.

Returning to the amplifier of Fig. 7.10, we find that replacing the MOSFET with the small-signal model of Fig. 7.13(b) and eliminating the dc sources results in the amplifier small-signal equivalent circuit shown in Fig. 7.14, from which the following expression for the voltage gain can be found:

$$A_v = \frac{v_{ds}}{v_{gs}} = -g_m(R_D \,\|\, r_o) \tag{7.39}$$

Thus, the finite output resistance r_o reduces the magnitude of the voltage gain.

Although the analysis above is performed on an NMOS transistor, the results, and the equivalent-circuit models of Fig. 7.13, apply equally well to PMOS devices, except for using $|V_{GS}|, |V_t|, |V_{OV}|$, and $|V_A|$ and replacing k_n with k_p.

The Transconductance g_m We shall now take a closer look at the MOSFET transconductance given by Eq. (7.32), which we rewrite with $k_n = k_n'(W/L)$ as follows:

$$g_m = k_n'(W/L)(V_{GS} - V_t) = k_n'(W/L)V_{OV} \tag{7.40}$$

This relationship indicates that g_m is proportional to the process transconductance parameter $k_n' = \mu_n C_{ox}$ and to the W/L ratio of the MOS transistor; hence to obtain relatively large transconductance the device must be short and wide. We also see that for a given device the transconductance is proportional to the overdrive voltage, $V_{OV} = V_{GS} - V_t$, the amount by which the bias voltage V_{GS} exceeds the threshold voltage V_t. Note, however, that increasing g_m by biasing the device at a larger V_{GS} has the disadvantage of reducing the allowable voltage signal swing at the drain.

We can find another useful expression for g_m by substituting $\sqrt{2I_D/(k_n'(W/L))}$ [from Eq. (7.25)] for V_{OV} in Eq. (7.40):

$$g_m = \sqrt{2k_n'}\sqrt{W/L}\sqrt{I_D} \tag{7.41}$$

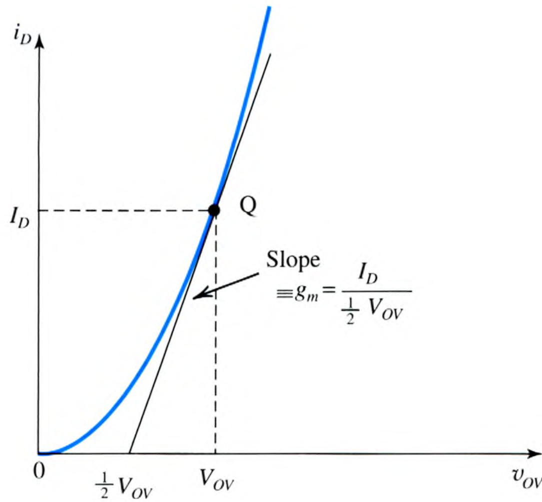


Figure 7.15 The slope of the tangent at the bias point Q intersects the v_{OV} axis at $\frac{1}{2}V_{OV}$. Thus, $g_m = I_D/(\frac{1}{2}V_{OV})$.

This expression shows us two things:

1. For a given MOSFET, g_m is proportional to the square root of the dc bias current.
2. At a given bias current, g_m is proportional to $\sqrt{W/L}$.

In contrast, as we will see shortly, the transconductance of the BJT is proportional to the bias current and is independent of the physical size and geometry of the device.

To gain some insight into the values of g_m obtained in MOSFETs, consider an integrated-circuit device operating at $I_D = 0.2$ mA and having $k_n' = 500\ \mu\text{A/V}^2$. Equation (7.41) shows that for $W/L = 1, g_m = 0.45$ mA/V, whereas a device for which $W/L = 100$ has $g_m = 4.5$ mA/V. In contrast, a BJT operating at a collector current of 0.2 mA has $g_m = 8$ mA/V.

Yet another useful expression for g_m of the MOSFET can be obtained by substituting for $k_n'(W/L)$ in Eq. (7.40) by $2I_D/(V_{GS} - V_t)^2$:

$$g_m = \frac{2I_D}{V_{GS} - V_t} = \frac{2I_D}{V_{OV}} = \frac{I_D}{(V_{OV}/2)} \tag{7.42}$$

A convenient graphical construction that clearly illustrates this relationship is shown in Fig. 7.15.

In summary, there are three different relationships for determining g_m—Eqs. (7.40), (7.41), and (7.42)—and there are three design parameters—(W/L), V_{OV}, and I_D, any two of which can be chosen independently. That is, the designer may choose to operate the MOSFET with a certain overdrive voltage V_{OV} and at a particular current I_D; the required W/L ratio can then be found and the resulting g_m determined.[3]

[3]This assumes that the circuit designer is also designing the device, as is typically the case in IC design. On the other hand, a circuit designer working with a discrete-circuit MOSFET obviously does not have the freedom to change its W/L ratio. Thus, in this case there are only two design parameters—V_{OV} and I_D, and only one can be specified by the designer.

EXERCISES

7.4 For the amplifier in Fig. 7.10, let $V_{DD} = 5$ V, $R_D = 10$ kΩ, $V_t = 1$ V, $k_n' = 20\ \mu\text{A/V}^2$, $W/L = 20$, $V_{GS} = 2$ V, and $\lambda = 0$.
(a) Find the dc current I_D and the dc voltage V_{DS}.
(b) Find g_m.
(c) Find the voltage gain.
(d) If $v_{gs} = 0.2 \sin \omega t$ volts, find v_{ds} assuming that the small-signal approximation holds. What are the minimum and maximum values of v_{DS}?
(e) Use Eq. (7.28) to determine the various components of i_D. Using the identity $\sin^2\omega t = \frac{1}{2} - \frac{1}{2}\cos 2\,\omega t$, show that there is a slight shift in I_D (by how much?) and that there is a second-harmonic component (i.e., a component with frequency $2\,\omega$). Express the amplitude of the second-harmonic component as a percentage of the amplitude of the fundamental. (This value is known as the second-harmonic distortion.)
Ans. (a) 0.2 mA, 3 V; (b) 0.4 mA/V; (c) –4 V/V; (d) $v_{ds} = -0.8 \sin\omega t$ volts, 2.2 V, 3.8 V; (e) $i_D = (204 + 80\ \sin \omega t - 4 \cos 2\,\omega t)\ \mu\text{A}$, 5%

7.5 An NMOS transistor has $\mu_n C_{ox} = 60\ \mu\text{A/V}^2$, $W/L = 40$, $V_t = 1$ V, and $V_A = 15$ V. Find g_m and r_o when (a) the bias voltage $V_{GS} = 1.5$ V, (b) the bias current $I_D = 0.5$ mA.
Ans. (a) 1.2 mA/V, 50 kΩ; (b) 1.55 mA/V, 30 kΩ

7.6 A MOSFET is to operate at $I_D = 0.1$ mA and is to have $g_m = 1$ mA/V. If $k_n' = 500\ \mu\text{A/V}^2$, find the required W/L ratio and the overdrive voltage.
Ans. 10; 0.2 V

7.7 For a fabrication process for which $\mu_p \simeq 0.4\,\mu_n$, find the ratio of the width of a PMOS transistor to the width of an NMOS transistor so that the two devices have equal g_m for the same bias conditions. The two devices have equal channel lengths.
Ans. 2.5

7.8 A PMOS transistor has $V_t = -1$ V, $k_p' = 60\ \mu\text{A/V}^2$, and $W/L = 16\ \mu\text{m}/0.8\ \mu\text{m}$. Find I_D and g_m when the device is biased at $V_{GS} = -1.6$ V. Also, find the value of r_o if λ (at $L = 1\ \mu$m) $= -0.04\ \text{V}^{-1}$.
Ans. 216 μA; 0.72 mA/V; 92.6 kΩ

7.9 Derive an expression for $(g_m r_o)$ in terms of V_A and V_{OV}. As we shall see in Chapter 8, this is an important transistor parameter and is known as the intrinsic gain. Evaluate the value of $g_m r_o$ for an NMOS transistor fabricated in a 0.18-μm CMOS process for which $V_A' = 6$ V/μm of channel length. Let the device have three times the minimum channel length and be operated at an overdrive voltage of 0.2 V.
Ans. $g_m r_o = 2V_A/V_{OV}$; 32.4 V/V

Example 7.3

Figure 7.16(a) shows a discrete MOSFET amplifier biased with a drain-to-gate resistance R_G. We will study this biasing arrangement in Section 7.4. The input signal v_i is coupled to the gate via a large capacitor, and the output signal at the drain is coupled to the load resistance R_L via another large

capacitor. We want to analyze this amplifier circuit to determine its small-signal voltage gain, its input resistance, and the largest allowable input signal. The transistor has $V_t = 1.5$ V, $k_n'(W/L) = 0.25$ mA/V^2, and $V_A = 50$ V. Assume that the coupling capacitors are large enough that they act as short circuits at the signal frequencies of interest.

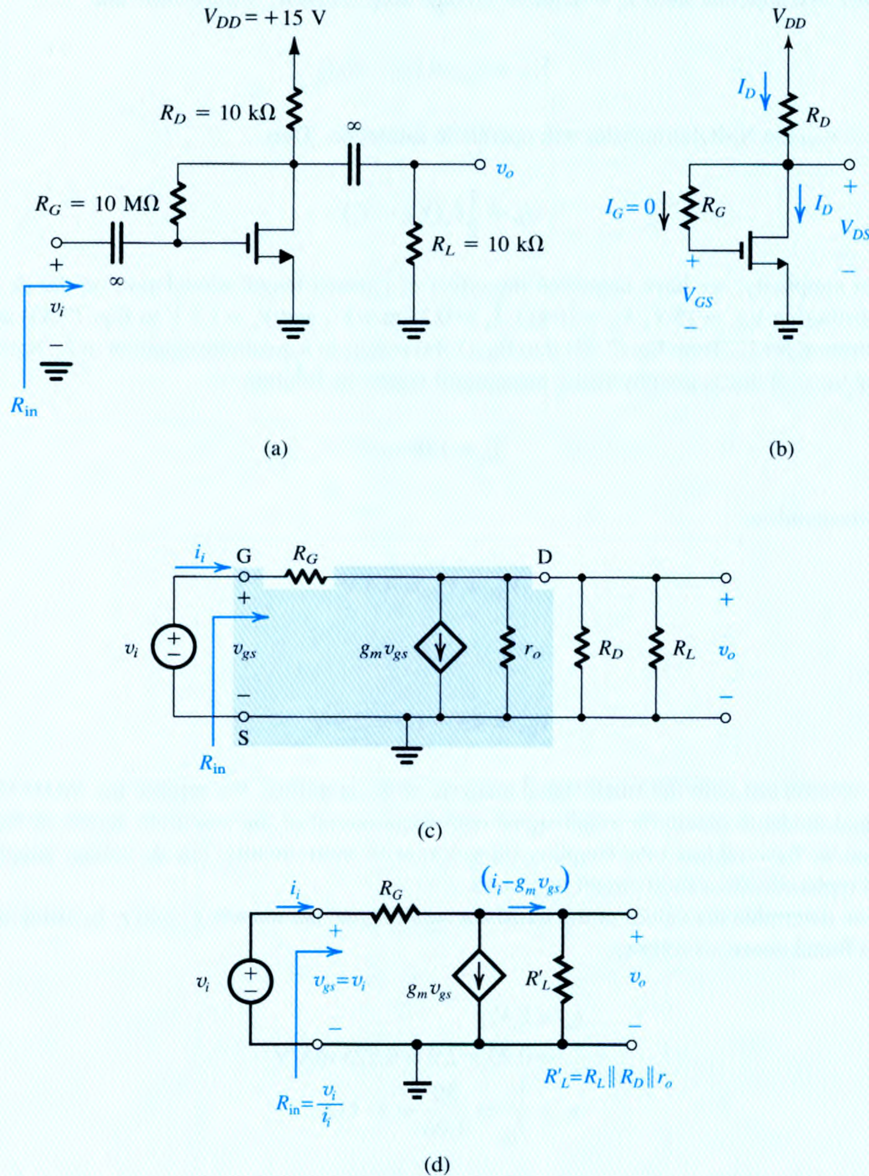


Figure 7.16 Example 7.3: **(a)** amplifier circuit; **(b)** circuit for determining the dc operating point; **(c)** the amplifier small-signal equivalent circuit; **(d)** a simplified version of the circuit in **(c)**.

Example 7.3 *continued*

Solution

We first determine the dc operating point. For this purpose, we eliminate the input signal v_i, and open-circuit the two coupling capacitors (since they block dc currents). The result is the circuit shown in Fig. 7.16(b). We note that since $I_G = 0$, the dc voltage drop across R_G will be zero, and

$$V_{GS} = V_{DS} = V_{DD} - R_D I_D \tag{7.43}$$

With $V_{DS} = V_{GS}$, the NMOS transistor will operate in saturation. Thus,

$$I_D = \frac{1}{2}k_n(V_{GS} - V_t)^2 \tag{7.44}$$

where, for simplicity, we have neglected the effect of channel-length modulation on the dc operating point. Substituting $V_{DD} = 15$ V, $R_D = 10$ kΩ, $k_n = 0.25$ mA/V², and $V_t = 1.5$ V in Eqs. (7.43) and (7.44), and substituting for V_{GS} from Eq. (7.43) into Eq. (7.44) results in a quadratic equation in I_D. Solving it and discarding the root that is not physically meaningful yields the solution

$$I_D = 1.06 \text{ mA}$$

which corresponds to

$$V_{GS} = V_{DS} = 4.4 \text{ V}$$

and

$$V_{OV} = 4.4 - 1.5 = 2.9 \text{ V}$$

Next we proceed with the small-signal analysis of the amplifier. We replace the MOSFET with its small-signal model to obtain the small-signal equivalent circuit of the amplifier, shown in Fig. 7.16(c). Notice that we have replaced the coupling capacitors with short circuits. The dc voltage supply V_{DD} has also been replaced with a short circuit to ground.

We can determine the values of the transistor small-signal parameters g_m and r_o by using the dc bias quantities found above, as follows:

$$\begin{aligned} g_m &= k_n V_{OV} \\ &= 0.25 \times 2.9 = 0.725 \text{ mA/V} \\ r_o &= \frac{V_A}{I_D} = \frac{50}{1.06} = 47 \text{ k}\Omega \end{aligned}$$

Next we use the equivalent circuit of Fig. 7.16(c) to determine the input resistance $R_{in} \equiv v_i/i_i$ and the voltage gain $A_v = v_o/v_i$. To that end we simplify the circuit by combining the three parallel resistances r_o,

R_D, and R_L in a single resistance R_L',

$$R_L' = R_L || R_D || r_o$$
$$= 10||10||47 = 4.52 \text{ k}\Omega$$

as shown in Fig. 7.16(d). For the latter circuit we can write the two equations

$$v_o = (i_i - g_m v_{gs})R_L' \quad (7.45)$$

and

$$i_i = \frac{v_{gs} - v_o}{R_G} \quad (7.46)$$

Substituting for i_i from Eq. (7.46) into Eq. (7.45) gives us the following expression for the voltage gain $A_v \equiv v_o/v_i = v_o/v_{gs}$:

$$A_v = -g_m R_L' \frac{1 - (1/g_m R_G)}{1 + (R_L'/R_G)}$$

Since R_G is very large, $g_m R_G \gg 1$ and $R_L'/R_G \ll 1$ (you can easily verify this), the gain expression can be approximated as

$$A_v \simeq -g_m R_L' \quad (7.47)$$

Substituting $g_m = 0.725$ mA/V and $R_L' = 4.52$ kΩ yields

$$A_v = -3.3 \text{ V/V}$$

To find the input resistance, we substitute in Eq. (7.46) for $v_o = A_v v_{gs} = -g_m R_L' v_{gs}$, then use $R_{in} \equiv v_i/i_i = v_{gs}/i_i$ to get

$$R_{in} = \frac{R_G}{1 + g_m R_L'} \quad (7.48)$$

This is an interesting relationship: The input resistance decreases as the gain $(g_m R_L')$ is increased. The value of R_{in} can now be determined; it is

$$R_{in} = \frac{10 \text{ M}\Omega}{1 + 3.3} = 2.33 \text{ M}\Omega$$

which is still very large.

The largest allowable input signal $\hat{v}_i$ is constrained by the need to keep the transistor in saturation at all times; that is,

$$v_{DS} \geq v_{GS} - V_t$$

Example 7.3 *continued*

We can enforce this condition by making the two terms equal at the point v_{GS} is maximum and v_{DS} is minimum,

$$v_{DS\min} = v_{GS\max} - V_t$$

$$V_{DS} - |A_v|\hat{v}_i = V_{GS} + \hat{v}_i - V_t$$

Since $V_{DS} = V_{GS}$, we obtain

$$\hat{v}_i = \frac{V_t}{|A_v| + 1}$$

This is a general relationship that applies to this circuit irrespective of the component values. Observe that it simply states that the maximum signal swing is determined by the fact that the bias arrangement makes $V_D = V_G$ and thus, to keep the MOSFET out of the triode region, the maximum signal between D and G is constrained to be equal to V_t. For our particular design,

$$\hat{v}_i = \frac{1.5}{3.3 + 1} = 0.35 \text{ V}$$

Since $V_{OV} = 2.9$ V, a v_i of 0.35 is indeed much smaller than $2V_{OV} = 5.8$ V; thus our assumption of linear operation is justified.

A modification of this circuit that increases the allowable signal swing is investigated in Problem 7.103.

EXERCISE

D7.10 Consider the amplifier circuit of Fig. 7.16(a) without the load resistance R_L and with channel-length modulation neglected. Let $V_{DD} = 5$ V, $V_t = 0.7$ V, and $k_n = 1$ mA/V^2. Find V_{OV}, I_D, R_D, and R_G to obtain a voltage gain of -25 V/V and an input resistance of 0.5 M Ω. What is the maximum allowable input signal, $\hat{v}_i$?

Ans. 0.319 V; 50.9 μA; 78.5 kΩ; 13 MΩ; 27 mV

Video Example VE 7.1 Analysis of a MOS Amplifier

In the circuit of Fig. VE7.1, the NMOS transistor has $|V_t| = 0.5$ V and $V_A = 50$ V and operates with $V_D =$ 1 V. What is the voltage gain v_o/v_i ? What do V_D and the gain become for I increased to 1 mA?

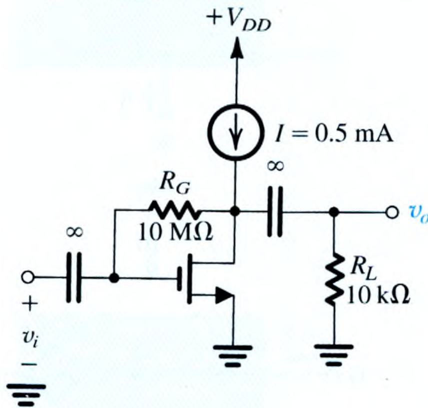


Figure VE7.1 Circuit for Video Example 7.1.

Solution: Go to **www.oup.com/he/sedra-smith8e** to watch the authors solve this problem.


VE 7.1

Related end-of-chapter problem: 7.31

The T Equivalent-Circuit Model Through a simple circuit transformation it is possible to develop an alternative equivalent-circuit model for the MOSFET. The development of such a model, known as the T model, is illustrated in Fig. 7.17. Figure 7.17(a) shows the equivalent circuit studied above without r_o. In Fig. 7.17(b) we have added a second $g_m v_{gs}$ current source in series with the original controlled source. This addition obviously does not change the terminal currents and is thus allowed. The newly created circuit node, labeled X, is joined to the gate terminal G in Fig. 7.17(c). Observe that the gate current does not change—that is, it remains equal to zero—and thus this connection does not alter the terminal characteristics. We now have a controlled current source $g_m v_{gs}$ connected across its control voltage v_{gs}. We can replace this controlled source by a resistance as long as this resistance draws an equal current as the source. (See the source-absorption theorem in Appendix D.) Thus the value of the resistance is $v_{gs}/g_m v_{gs} = 1/g_m$. This replacement is shown in Fig. 7.17(d), which depicts the alternative model. Notice that i_g is still zero, $i_d = g_m v_{gs}$, and $i_s = v_{gs}/(1/g_m) = g_m v_{gs}$, all the same as in the original model in Fig. 7.17(a).

The model of Fig. 7.17(d) shows that the resistance between gate and source looking into the source is $1/g_m$. This observation and the T model prove useful in many applications. Note that the resistance between gate and source, looking into the gate, is infinite.

In developing the T model we did not include r_o. If we want, we can do this by incorporating in the circuit of Fig. 7.17(d) a resistance r_o between drain and source, as shown in Fig. 7.18(a). An alternative representation of the T model, in which the voltage-controlled current source is replaced with a current-controlled current source, is shown in Fig. 7.18(b).

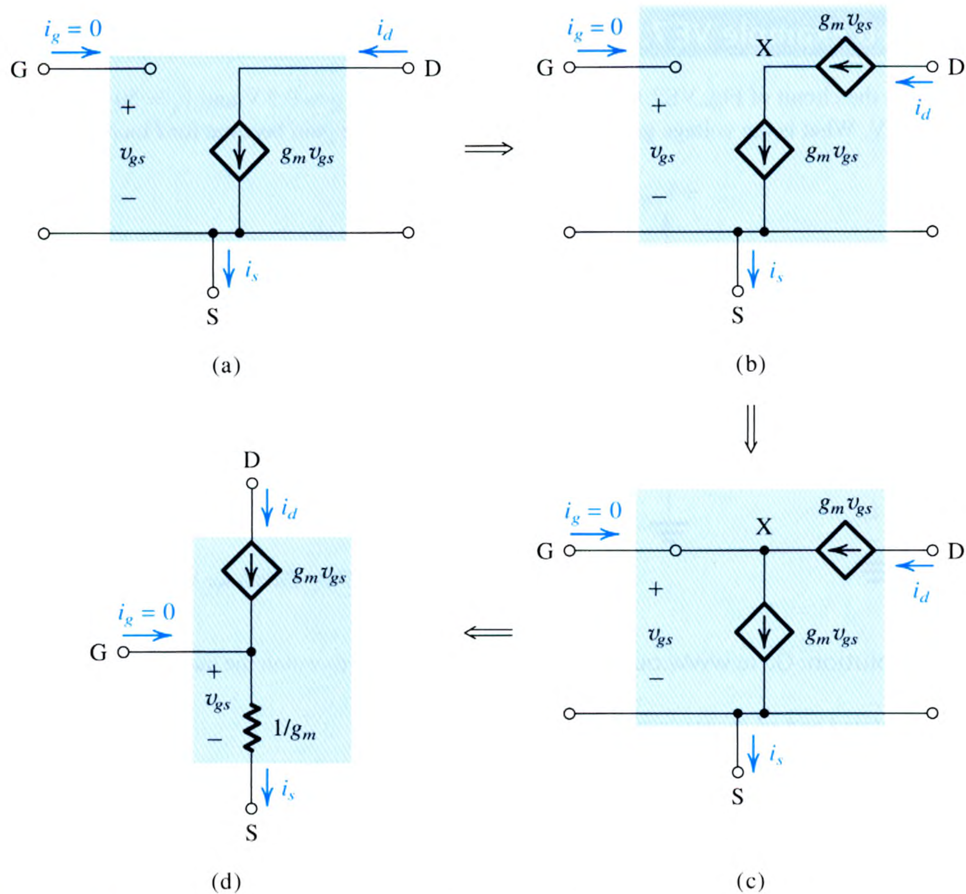


Figure 7.17 Development of the T equivalent-circuit model for the MOSFET. For simplicity, r_o has been omitted; however, it may be added between D and S in the T model of **(d)**.

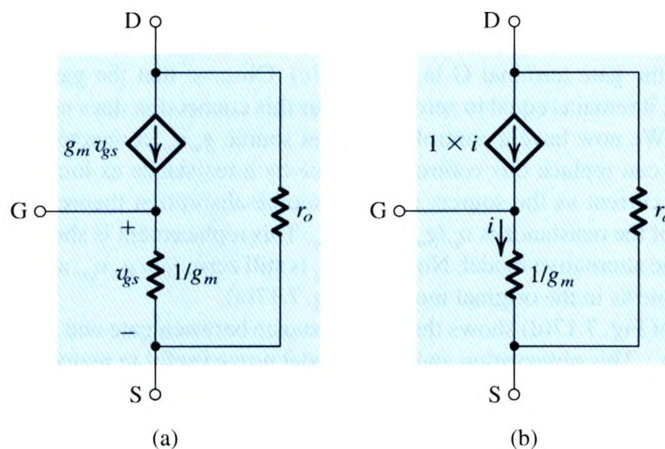


Figure 7.18 **(a)** The T model of the MOSFET augmented with the drain-to-source resistance r_o. **(b)** An alternative representation of the T model.

Finally, note that in order to distinguish the model of Fig. 7.13(b) from the equivalent T model, the former is sometimes referred to as the **hybrid-π model**, a carryover from the bipolar transistor literature. We will explain the origin of this name shortly.

Example 7.4

Figure 7.19(a) shows a MOSFET amplifier biased by a constant-current source I. Assume that the values of I and R_D are such that the MOSFET operates in the saturation region. The input signal v_i is coupled to the source terminal by a large capacitor C_{C1}. Similarly, the output signal at the drain is taken through a large coupling capacitor C_{C2}. Find the input resistance R_{in} and the voltage gain v_o/v_i. Neglect channel-length modulation.

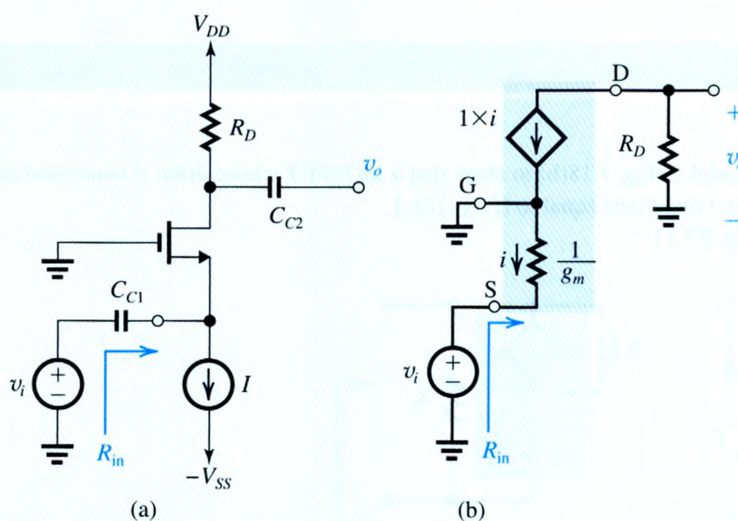


Figure 7.19 **(a)** Amplifier circuit for Example 7.4. **(b)** Small-signal equivalent circuit of the amplifier in **(a)**.

Solution

Replacing the MOSFET with its T equivalent-circuit model results in the amplifier equivalent circuit shown in Fig. 7.19(b). Notice that the dc current source I is replaced with an open circuit and the dc voltage source V_{DD} is replaced by a short circuit. The large coupling capacitors have been replaced by short circuits. From the equivalent-circuit model we determine

$$R_{in} = \frac{v_i}{-i} = 1/g_m$$

Example 7.4 *continued*

and

$$v_o = -iR_D = \left(\frac{v_i}{1/g_m}\right)R_D = g_m R_D v_i$$

Thus,

$$A_v \equiv \frac{v_o}{v_i} = g_m R_D$$

This amplifier, known as the common-gate amplifier because the gate at ground potential is common to both the input and output ports, has a low input resistance $(1/g_m)$ and a noninverting gain. We will study this amplifier type in Section 7.3.5.

EXERCISE

7.11 Use the T model of Fig. 7.18(b) to show that a MOSFET whose drain is connected to its gate exhibits an incremental resistance equal to $[(1/g_m) \parallel r_o]$.
Ans. See Fig. E7.11.

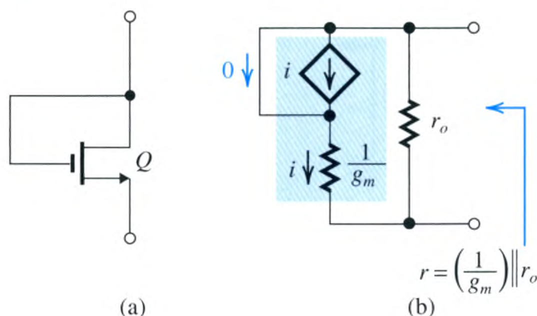


Figure E7.11 Circuits for Exercise 7.11. Note that the bias arrangement of Q is not shown.

Modeling the Body Effect As we mentioned earlier (see Section 5.4), the body effect occurs in a MOSFET when the source is not tied to the substrate (which is always connected to the most negative power supply in the integrated circuit for n-channel devices and to the most positive for p-channel devices). Thus the substrate (body) will be at signal ground, but since the source is not, a signal voltage v_{bs} develops between the body (B) and the source (S). The substrate then acts as a "second gate" or a **backgate** for the MOSFET. Thus the signal v_{bs} gives rise to a drain-current component, which we will write as $g_{mb} v_{bs}$, where g_{mb} is the

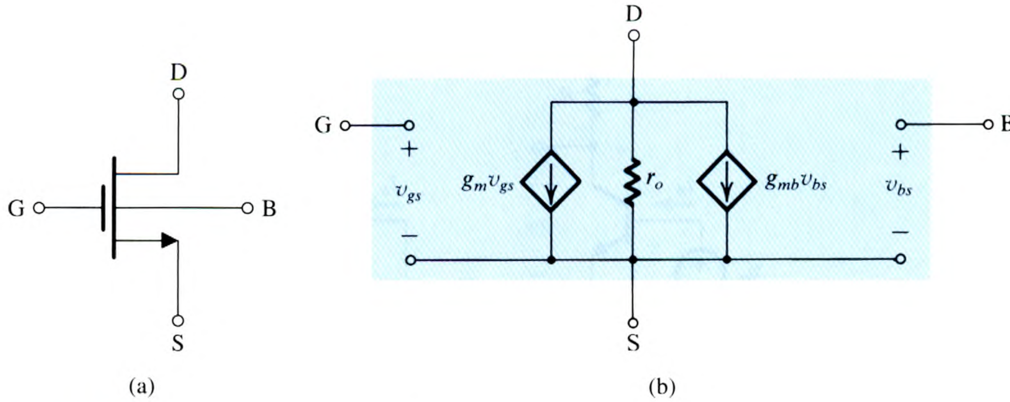


Figure 7.20 Small-signal, equivalent-circuit model of a MOSFET in which the source is not connected to the body.

body transconductance, defined as

$$g_{mb} \equiv \left.\frac{\partial i_D}{\partial v_{BS}}\right|_{\substack{v_{GS}=\text{constant} \\ v_{DS}=\text{constant}}} \tag{7.49}$$

Recalling that i_D depends on v_{BS} through the dependence of V_t on V_{BS}, we can show that

$$g_{mb} = \chi g_m \tag{7.50}$$

where

$$\chi \equiv \frac{\partial V_t}{\partial V_{SB}} = \frac{\gamma}{2\sqrt{2\phi_f + V_{SB}}} \tag{7.51}$$

Typically the value of χ lies in the range 0.1 to 0.3.

Figure 7.20 shows the MOSFET model augmented to include the controlled source $g_{mb}v_{bs}$ that models the body effect. Ideally, this is the model to be used whenever the source is not connected to the substrate. However, except in some very particular situations, the body effect can generally be ignored in the initial, pencil-and-paper design of MOSFET amplifiers.

Finally, although the analysis above was performed on an NMOS transistor, the results and the equivalent circuit of Fig. 7.20 apply equally well to PMOS transistors, except for using $|V_{GS}|$, $|V_t|$, $|V_{OV}|$, $|V_A|$, $|V_{SB}|$, $|\gamma|$, and $|\lambda|$ and replacing k_n' with k_p' in the appropriate formula.

7.2.2 The BJT Case

We next consider the small-signal operation of the BJT and develop small-signal equivalent-circuit models that represent its operation at a given bias point. The following development parallels what we used for the MOSFET except that here we have an added complication: The BJT draws a finite base current. As we will see shortly, this phenomenon (finite β) manifests itself as a finite input resistance looking into the base of the BJT (as compared to the infinite input resistance looking into the gate of the MOSFET).

Consider the *conceptual* amplifier circuit shown in Fig. 7.21(a). Here the base–emitter junction is forward biased by a dc voltage V_{BE}. The reverse bias of the collector–base junction

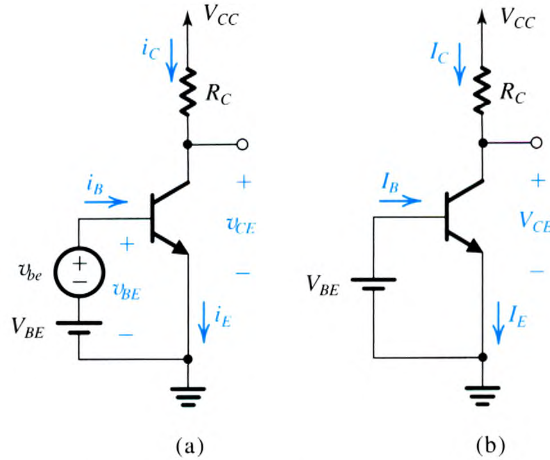


Figure 7.21 (**a**) Conceptual circuit to illustrate the operation of the transistor as an amplifier. (**b**) The circuit of (**a**) with the signal source v_{be} eliminated for dc (bias) analysis.

is established by connecting the collector to another power supply of voltage V_{CC} through a resistor R_C. The input signal to be amplified is represented by the voltage source v_{be} that is superimposed on V_{BE}.

The DC Bias Point We first consider the dc bias conditions by setting the signal v_{be} to zero. The circuit reduces to that in Fig. 7.21(b), and we can write the following relationships for the dc currents and voltages:

$$I_C = I_S e^{V_{BE}/V_T} \tag{7.52}$$

$$I_E = I_C/\alpha \tag{7.53}$$

$$I_B = I_C/\beta \tag{7.54}$$

$$V_{CE} = V_{CC} - I_C R_C \tag{7.55}$$

For active-mode operation, V_{CE} should be greater than $(V_{BE} - 0.4)$ by an amount that allows for the required negative signal swing at the collector.

The Collector Current and the Transconductance If a signal v_{be} is applied as shown in Fig. 7.21(a), the total instantaneous base–emitter voltage v_{BE} becomes

$$v_{BE} = V_{BE} + v_{be}$$

Correspondingly, the collector current becomes

$$i_C = I_S e^{v_{BE}/V_T} = I_S e^{(V_{BE}+v_{be})/V_T}$$
$$= I_S e^{V_{BE}/V_T} e^{v_{be}/V_T}$$

Use of Eq. (7.52) yields

$$i_C = I_C e^{v_{be}/V_T} \tag{7.56}$$

Now, if $v_{be} \ll V_T$, we can approximate Eq. (7.56) as

$$i_C \simeq I_C\left(1 + \frac{v_{be}}{V_T}\right) \tag{7.57}$$

Here we have expanded the exponential in Eq. (7.56) in a series and retained only the first two terms. That is, we have assumed that

$$v_{be} \ll V_T \tag{7.58}$$

so that we can neglect the higher-order terms in the exponential series expansion. The condition in Eq. (7.58) is the **small-signal approximation** for the BJT and corresponds to that in Eq. (7.29) for the MOSFET case. The small-signal approximation for the BJT is valid only for v_{be} less than 5 mV, or 10 mV at most. Under this approximation, the total collector current is given by Eq. (7.57) and can be rewritten

$$i_C = I_C + \frac{I_C}{V_T}v_{be} \tag{7.59}$$

Thus the collector current is composed of the dc bias value I_C and a signal component i_c,

$$i_c = \frac{I_C}{V_T}v_{be} \tag{7.60}$$

This equation relates the signal current in the collector to the corresponding base–emitter signal voltage. We can rewrite it as

$$i_c = g_m v_{be} \tag{7.61}$$

where g_m is the **transconductance**, and, from Eq. (7.60), it is given by

$$g_m = \frac{I_C}{V_T} \tag{7.62}$$

We observe that the transconductance of the BJT is directly proportional to the collector bias current I_C. Thus to obtain a constant predictable value for g_m, we need a constant predictable I_C. Also, we note that BJTs have relatively high transconductance in comparison to MOSFETs; for instance, at $I_C = 1$ mA, $g_m \simeq 40$ mA/V. Finally, unlike the MOSFET, whose g_m depends on the device dimensions (W and L), g_m of a BJT depends only on the dc collector current at which it is biased to operate.

A graphical representation of g_m is given in Fig. 7.22, where g_m is equal to the slope of the tangent to the i_C–v_{BE} characteristic curve at $i_C = I_C$ (i.e., at the bias point Q). Thus,

$$g_m = \left.\frac{\partial i_C}{\partial v_{BE}}\right|_{i_C = I_C} \tag{7.63}$$

The small-signal approximation implies keeping the signal amplitude small enough that *operation is restricted to an almost-linear segment of the i_C–v_{BE} exponential curve.* If we increase the signal amplitude, the collector current will have components nonlinearly related to v_{be}.

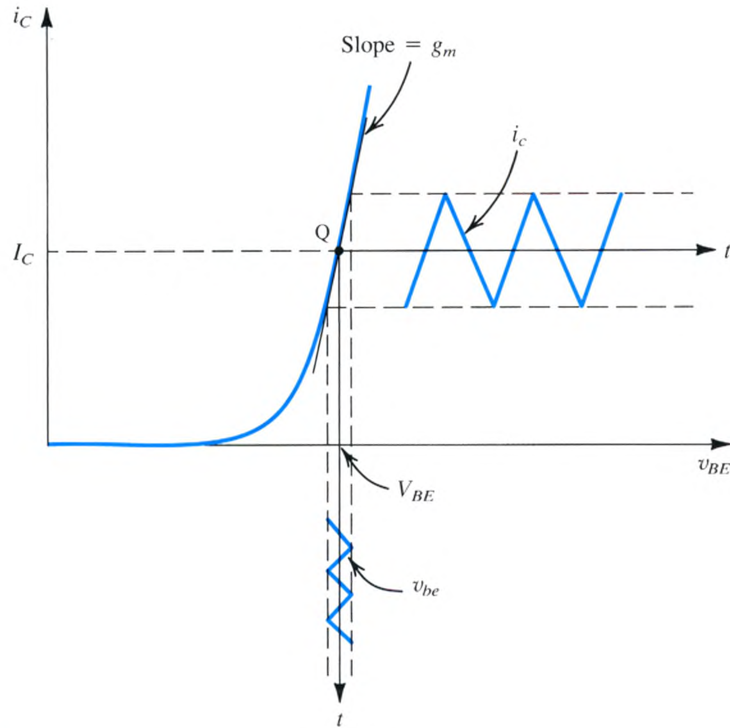


Figure 7.22 Linear operation of the BJT under the small-signal condition: A small-signal v_{be} with a triangular waveform is superimposed on the dc voltage V_{BE}. It gives rise to a collector-signal current i_c, also of triangular waveform, superimposed on the dc current I_C. Here, $i_c = g_m v_{be}$, where g_m is the slope of the i_C–v_{BE} curve at the bias point Q.

EXERCISES

7.12 Use Eq. (7.63) to derive the expression for g_m in Eq. (7.62).

7.13 Calculate the value of g_m for a BJT biased at $I_C = 0.5$ mA.
Ans. 20 mA/V

The Base Current and the Input Resistance at the Base To determine the resistance seen by v_{be}, we first evaluate the total base current i_B using Eq. (7.59), as follows:

$$i_B = \frac{i_C}{\beta} = \frac{I_C}{\beta} + \frac{1}{\beta}\frac{I_C}{V_T}v_{be}$$

Thus,

$$i_B = I_B + i_b \tag{7.64}$$

where I_B is equal to I_C/β and the signal component i_b is given by

$$i_b = \frac{1}{\beta}\frac{I_C}{V_T}v_{be} \tag{7.65}$$

Substituting for I_C/V_T by g_m gives

$$i_b = \frac{g_m}{\beta}v_{be} \tag{7.66}$$

The small-signal input resistance between base and emitter, *looking into the base,* is denoted by r_π and is defined as

$$r_\pi \equiv \frac{v_{be}}{i_b} \tag{7.67}$$

Using Eq. (7.66) gives

$$r_\pi = \frac{\beta}{g_m} \tag{7.68}$$

Thus r_π is directly proportional to β and is inversely proportional to g_m and thus to the bias current I_C. Substituting for g_m in Eq. (7.68) from Eq. (7.62) and replacing I_C/β by I_B gives an alternative expression for r_π,

$$r_\pi = \frac{V_T}{I_B} \tag{7.69}$$

Recall that since the gate current of the MOSFET is zero (at dc and low frequencies) the input resistance at the gate is infinite; that is, in the MOSFET there is no counterpart to r_π.[4]

EXERCISE

7.14 A BJT amplifier is biased to operate at a constant collector current $I_C = 0.5$ mA irrespective of the value of β. If the manufacturer specifies β to range from 50 to 200, give the expected range of g_m, I_B, and r_π.

Ans. g_m is constant at 20 mA/V; $I_B = 10\ \mu$A to 2.5 μA; $r_\pi = 2.5$ kΩ to 10 kΩ

The Emitter Current and the Input Resistance at the Emitter We can find the total emitter current i_E using Eq. (7.59) as

$$i_E = \frac{i_C}{\alpha} = \frac{I_C}{\alpha} + \frac{i_c}{\alpha}$$

[4] At high frequencies, the input capacitance at the MOSFET gate makes the input current finite (see Chapter 10).

Thus,

$$i_E = I_E + i_e \tag{7.70}$$

where I_E is equal to I_C/α and the signal current i_e is given by

$$i_e = \frac{i_c}{\alpha} = \frac{I_C}{\alpha V_T} v_{be} = \frac{I_E}{V_T} v_{be} \tag{7.71}$$

If we denote the small-signal resistance between base and emitter *looking into the emitter* by r_e, it can be defined as

$$r_e \equiv \frac{v_{be}}{i_e} \tag{7.72}$$

Using Eq. (7.71) we find that r_e, called the **emitter resistance**, is given by

$$r_e = \frac{V_T}{I_E} \tag{7.73}$$

Comparison with Eq. (7.62) reveals that

$$r_e = \frac{\alpha}{g_m} \simeq \frac{1}{g_m} \tag{7.74}$$

We can find the relationship between r_π and r_e by combining their respective definitions in Eqs. (7.67) and (7.72) as

$$v_{be} = i_b r_\pi = i_e r_e$$

Thus,

$$r_\pi = (i_e/i_b) r_e$$

which yields

$$r_\pi = (\beta + 1) r_e \tag{7.75}$$

Figure 7.23 illustrates the definition of r_π and r_e.

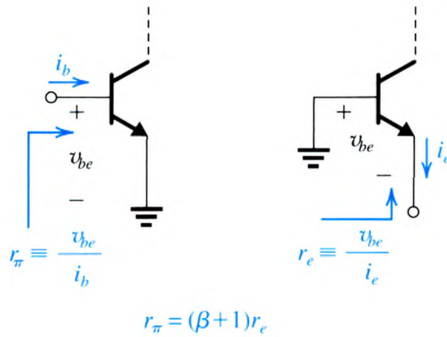


Figure 7.23 Illustrating the definition of r_π and r_e.

Finally, a comparison with the MOSFET would be useful: For the MOSFET, $\alpha = 1$ and the resistance looking into the source is simply $1/g_m$.

SHOCKLEY AND SILICON VALLEY

In 1956 William Bradford Shockley started a new company, Shockley Semiconductor Laboratory, in Mountain View, California (near Stanford, his birthplace). While at Bell Labs, together with John Bardeen and Walter Brattain, he had invented the BJT. At Shockley, he concentrated on developing semiconductor devices, particularly a new four-layer diode. It was a propitious time: The first *Sputnik* was launched a year later, and the ensuing space race accelerated demand for solid-state circuits. But Shockley's scientific genius and ability to select and attract good team members, first demonstrated at Bell Labs, was not accompanied by comparable talent for management, and in 1957, eight of his team members (the so-called Traitorous Eight, including Gordon Moore and Robert Noyce) left Shockley to create Fairchild Semiconductor. Nevertheless, Shockley is the man who brought silicon to Silicon Valley! In his later years, Shockley became interested in eugenics and made many controversial and abhorrent statements about the relationship between race and intelligence. This ultimately destroyed his reputation.

EXERCISE

7.15 A BJT having $\beta = 100$ is biased at a dc collector current of 1 mA. Find the value of g_m, r_e, and r_π at the bias point.
Ans. 40 mA/V; 25 Ω; 2.5 kΩ

The Voltage Gain The total collector voltage v_{CE} is

$$\begin{aligned} v_{CE} &= V_{CC} - i_C R_C \\ &= V_{CC} - (I_C + i_c) R_C \\ &= (V_{CC} - I_C R_C) - i_c R_C \\ &= V_{CE} - i_c R_C \end{aligned} \quad (7.76)$$

Thus, superimposed on the collector bias voltage V_{CE} we have signal voltage v_{ce} given by

$$\begin{aligned} v_{ce} &= -i_c R_C = -g_m v_{be} R_C \\ &= (-g_m R_C) v_{be} \end{aligned} \quad (7.77)$$

from which we find the voltage gain A_v of this amplifier as

$$A_v \equiv \frac{v_{ce}}{v_{be}} = -g_m R_C \quad (7.78)$$

Here again we note that because g_m is directly proportional to the collector bias current, the gain will be as stable as the collector bias current. Substituting for g_m from Eq. (7.62) enables

us to express the gain in the form

$$A_v = -\frac{I_C R_C}{V_T} \tag{7.79}$$

which is identical to the expression we derived in Section 7.1 (Eq. 7.21). Finally, we note that the gain expression in Eq. (7.78) is identical in form to that for the MOSFET amplifier (namely, $-g_m R_D$).

EXERCISE

7.16 In the circuit of Fig. 7.21(a), V_{BE} is adjusted to yield a dc collector current of 1 mA. Let $V_{CC} = 15$ V, $R_C = 10\ \text{k}\Omega$, and $\beta = 100$. Find the voltage gain v_{ce}/v_{be}. If $v_{be} = 0.005 \sin \omega t$ volt, find $v_C(t)$ and $i_B(t)$.
Ans. -400 V/V; $5 - 2 \sin \omega t$ volts; $10 + 2 \sin \omega t$ μA

Separating the Signal and the DC Quantities The analysis above indicates that every current and voltage in the amplifier circuit of Fig. 7.21(a) is composed of two components: a dc component and a signal component. For instance, $v_{BE} = V_{BE} + v_{be}$, $i_C = I_C + i_c$, and so on. The dc components are determined from the dc circuit given in Fig. 7.21(b) and from the relationships imposed by the transistor (Eqs. 7.52 through 7.54). On the other hand, we can represent the signal operation of the BJT by eliminating the dc sources, as shown in Fig. 7.24. Observe that since the voltage of an ideal dc supply does not change, the signal voltage across it will be zero. For this reason we have replaced V_{CC} and V_{BE} with short circuits. Had the circuit contained ideal dc current sources, these would have been replaced by open circuits. Note, however, that the circuit of Fig. 7.24 is useful only insofar as it shows the various signal currents and voltages; it is *not* an actual amplifier circuit, since the dc bias circuit is not shown.

Figure 7.24 also shows the expressions for the current increments (i_c, i_b, and i_e) obtained when a small signal v_{be} is applied. These relationships can be represented by a circuit. Such

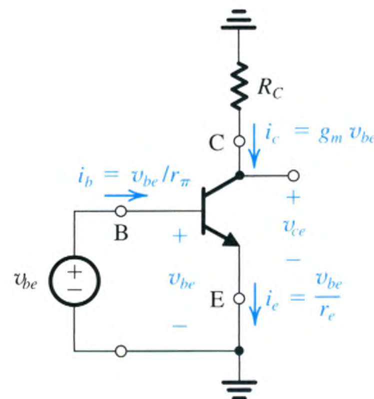


Figure 7.24 The amplifier circuit of Fig. 7.21(a) with the dc sources (V_{BE} and V_{CC}) eliminated (short-circuited). Thus only the signal components are present. Note that this is a representation of the signal operation of the BJT and not an actual amplifier circuit.

a circuit must have three terminals—C, B, and E—and must yield the same terminal currents indicated in Fig. 7.24. The resulting circuit is then *equivalent to the transistor as far as small-signal operation is concerned,* and thus it can be considered an equivalent small-signal circuit model.

The Hybrid-π Model An equivalent-circuit model for the BJT is shown in Fig. 7.25(a). This model represents the BJT as a voltage-controlled current source and explicitly includes the input resistance looking into the base, r_π. The model obviously yields $i_c = g_m v_{be}$ and $i_b = v_{be}/r_\pi$. Not so obvious, however, is the fact that the model also yields the correct expression for i_e. This can be shown as follows: At the emitter node we have

$$\begin{aligned} i_e &= \frac{v_{be}}{r_\pi} + g_m v_{be} = \frac{v_{be}}{r_\pi}(1 + g_m r_\pi) \\ &= \frac{v_{be}}{r_\pi}(1+\beta) = v_{be} \Big/ \left(\frac{r_\pi}{1+\beta}\right) \\ &= v_{be}/r_e \end{aligned}$$

A slightly different equivalent-circuit model can be obtained by expressing the current of the controlled source ($g_m v_{be}$) in terms of the base current i_b as follows:

$$\begin{aligned} g_m v_{be} &= g_m(i_b r_\pi) \\ &= (g_m r_\pi) i_b = \beta i_b \end{aligned}$$

This results in the alternative equivalent-circuit model shown in Fig. 7.25(b). Here the transistor is represented as a current-controlled current source, with the control current being i_b.

As we have done in the case of the MOSFET's small-signal models, we can account for the Early effect (the slight dependence of i_C on v_{CE} due to basewidth modulation) by adding the resistance $r_o = V_A/I_C$ between collector and emitter, as shown in Fig. 7.26. Note that to conform with the literature, we have renamed v_{be} as v_π. The two models of Fig. 7.26 are versions of the hybrid-π model, the most widely used model for the BJT. The equivalent circuit of Fig. 7.26(a) corresponds to that of the MOSFET [Fig. 7.13(b)] except for r_π, which

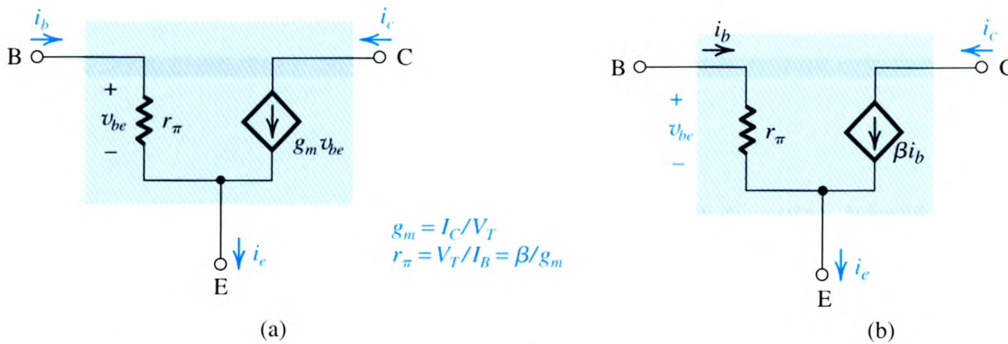


Figure 7.25 Two slightly different versions of the hybrid-π model for the small-signal operation of the BJT. The equivalent circuit in (**a**) represents the BJT as a voltage-controlled current source (a transconductance amplifier), and that in (**b**) represents the BJT as a current-controlled current source (a current amplifier).

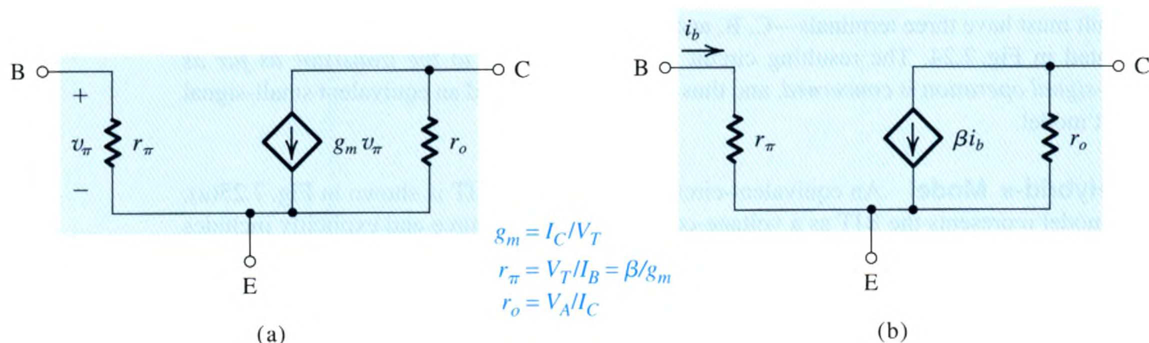


Figure 7.26 The hybrid-π small-signal model, in its two versions, with the resistance r_o included.

accounts for the finite base current (or finite β) of the BJT. However, the equivalent circuit of Fig. 7.26(b) has no MOS counterpart.

It is important to note that the small-signal equivalent circuits of Fig. 7.26 model the operation of the BJT *at a given bias point*. This should be obvious from the fact that the model parameters g_m, r_π, and r_o depend on the value of the dc bias current I_C, as indicated in Fig. 7.26. That is, these equivalent circuits model the *incremental operation* of the BJT around the bias point.

As in the case of the MOSFET amplifier, including r_o in the BJT model causes the voltage gain of the conceptual amplifier of Fig. 7.21(a) to become

$$\frac{v_o}{v_{be}} = -g_m(R_C \| r_o) \tag{7.80}$$

Thus, the magnitude of the gain is reduced somewhat.

EXERCISE

7.17 For the model in Fig. 7.25(b) show that $i_c = g_m v_{be}$ and $i_e = v_{be}/r_e$.

The T Model Although the hybrid-π model (in one of its two variants shown in Fig. 7.25) can be used to carry out small-signal analysis of any transistor circuit, there are situations in which an alternative model, shown in two versions in Fig. 7.27, is much more convenient. This model is called, as in the case of the MOSFET, the **T model**. The model of Fig. 7.27(a) represents the BJT as a voltage-controlled current source with the control voltage being v_{be}. Here, however, the resistance between base and emitter, looking into the emitter, is explicitly shown. From Fig. 7.27(a) we see clearly that the model yields the correct expressions for i_c and i_e. It can also be shown to yield the correct expression for i_b (see Exercise 7.18).

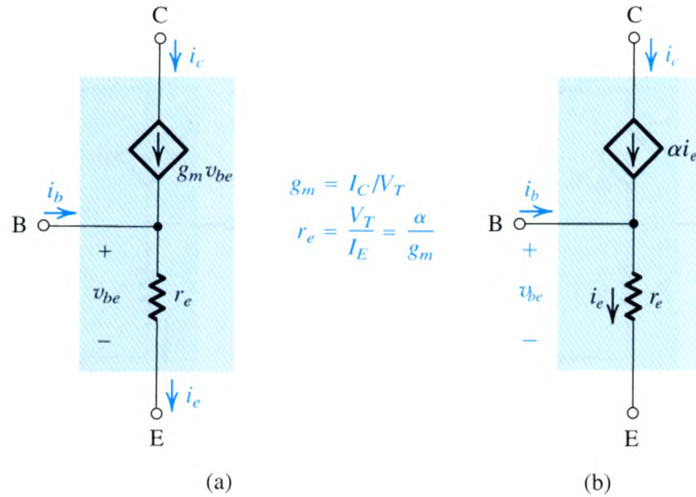


Figure 7.27 Two slightly different versions of what is known as the *T model* of the BJT. The circuit in (**a**) is a voltage-controlled current source representation and that in (**b**) is a current-controlled current source representation. These models explicitly show the emitter resistance r_e rather than the base resistance r_π featured in the hybrid-π model.

If in the model of Fig. 7.27(a) we express the current of the controlled source in terms of the emitter current as

$$\begin{aligned} g_m v_{be} &= g_m(i_e r_e) \\ &= (g_m r_e) i_e = \alpha i_e \end{aligned}$$

we obtain the alternative T model shown in Fig. 7.27(b). Here the BJT is represented as a current-controlled current source but with the control signal being i_e.

Finally, the T models can be augmented by r_o to account for the dependence of i_c to v_{ce} (the Early effect) to obtain the equivalent circuits shown in Fig. 7.28.

EXERCISE

7.18 Show that for the T model in Fig. 7.27(a), $i_b = v_{be}/r_\pi$.

Small-Signal Models of the *pnp* Transistor Although the small-signal models in Figs. 7.26 and 7.28 were developed for the case of the *npn* transistor, they apply equally well to the *pnp* transistor *with no change in polarities.*

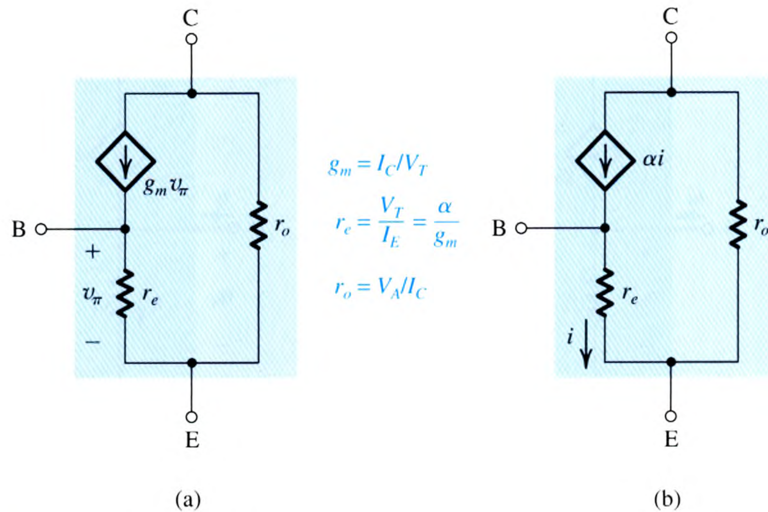


Figure 7.28 The T models of the BJT.

Example 7.5

We want to analyze the transistor amplifier shown in Fig. 7.29(a) to find its voltage gain v_o/v_i. Assume $\beta = 100$ and neglect the Early effect.

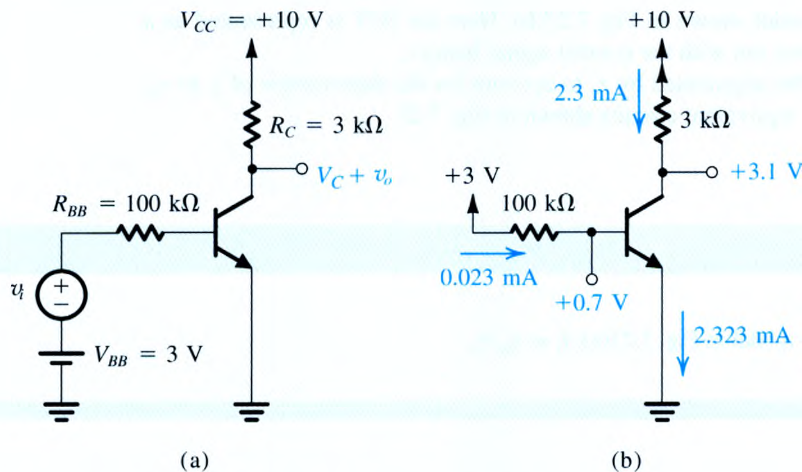


Figure 7.29 Example 7.5: (**a**) amplifier circuit; (**b**) circuit for dc analysis; (**c**) amplifier circuit with dc sources replaced by short circuits; (**d**) amplifier circuit with the transistor replaced by its hybrid-π, small-signal model.

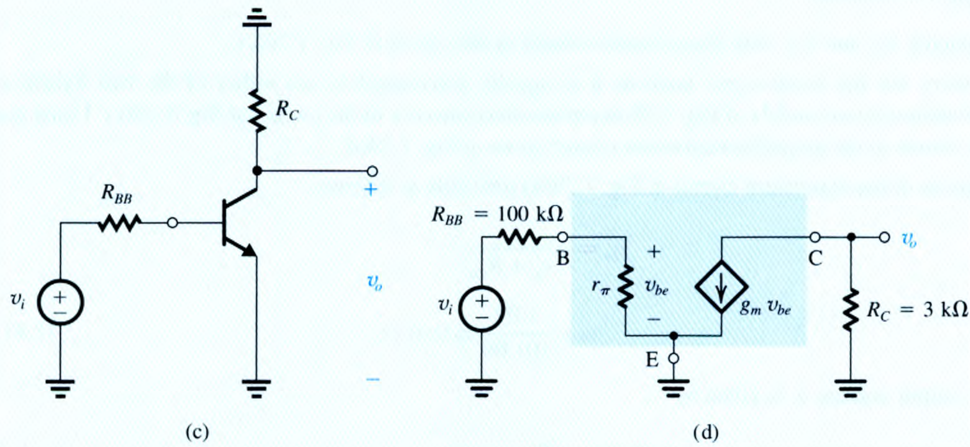


Figure 7.29 *continued*

Solution

We will follow a five-step process:

1. First, we need to determine the quiescent operating point. For this purpose we assume that $v_i = 0$ and thus obtain the dc circuit shown in Fig. 7.29(b). The dc base current will be

$$I_B = \frac{V_{BB} - V_{BE}}{R_{BB}}$$
$$\simeq \frac{3 - 0.7}{100} = 0.023 \text{ mA}$$

The dc collector current will be

$$I_C = \beta I_B = 100 \times 0.023 = 2.3 \text{ mA}$$

The dc voltage at the collector will be

$$V_C = V_{CC} - I_C R_C$$
$$= +10 - 2.3 \times 3 = +3.1 \text{ V}$$

Since V_B at +0.7 V is less than V_C, it follows that in the quiescent condition the transistor will be operating in the active mode. The dc analysis is illustrated in Fig. 7.29(b).

2. Having determined the operating point, we can now determine the small-signal model parameters:

$$r_e = \frac{V_T}{I_E} = \frac{25 \text{ mV}}{(2.3/0.99) \text{ mA}} = 10.8 \ \Omega$$
$$g_m = \frac{I_C}{V_T} = \frac{2.3 \text{ mA}}{0.025 \text{ V}} = 92 \text{ mA/V}$$
$$r_\pi = \frac{\beta}{g_m} = \frac{100}{92} = 1.09 \text{ k}\Omega$$

Example 7.5 *continued*

3. Replacing V_{BB} and V_{CC} with short circuits results in the circuit in Fig. 7.29(c).
4. To carry out the small-signal analysis it is equally convenient to use either of the two hybrid-π equivalent-circuit models of Fig. 7.25 to replace the transistor in the circuit of Fig. 7.29(c). Using the first results in the amplifier equivalent circuit given in Fig. 7.29(d).
5. Analysis of the equivalent circuit in Fig. 7.29(d) proceeds as follows:

$$v_{be} = v_i \frac{r_\pi}{r_\pi + R_{BB}}$$

$$= v_i \frac{1.09}{101.09} = 0.011 v_i \quad (7.81)$$

The output voltage v_o is given by

$$v_o = -g_m v_{be} R_C$$
$$= -92 \times 0.011 v_i \times 3 = -3.04 v_i$$

Thus the voltage gain will be

$$A_v = \frac{v_o}{v_i} = -3.04 \text{ V/V} \quad (7.82)$$

Example 7.6

To learn more about the way transistor amplifiers operate, we can consider the waveforms at various points in the circuit analyzed in the previous example. For this purpose assume that v_i has a triangular waveform. First determine the maximum amplitude that v_i is allowed to have. Then, with the amplitude of v_i set to this value, give the waveforms of the total quantities $i_B(t)$, $v_{BE}(t)$, $i_C(t)$, and $v_C(t)$.

Solution

One constraint on signal amplitude is the small-signal approximation, which stipulates that v_{be} should not exceed about 10 mV. If we take the triangular waveform v_{be} to be 20 mV peak-to-peak and work backward, Eq. (7.81) can be used to determine the maximum possible peak of v_i,

$$\hat{v}_i = \frac{\hat{v}_{be}}{0.011} = \frac{10}{0.011} = 0.91 \text{ V}$$

To check whether the transistor remains in the active mode with v_i having a peak value $\hat{v}_i = 0.91$V, we have to evaluate the collector voltage. The voltage at the collector will consist of a triangular wave v_o

superimposed on the dc value $V_C = 3.1$ V. The peak voltage of the triangular waveform will be

$$\hat{v}_o = \hat{v}_i \times \text{gain} = 0.91 \times 3.04 = 2.77 \text{ V}$$

It follows that when the output swings negative, the collector voltage reaches a minimum of $3.1 - 2.77 = 0.33$ V, which is lower than the base voltage by less than 0.4 V. Thus the transistor will remain in the active mode with v_i having a peak value of 0.91 V. Nevertheless, to be on the safe side, we will use a somewhat lower value for $\hat{v}_i$ of approximately 0.8 V, as shown in Fig. 7.30(a), and complete the analysis of this problem utilizing the equivalent circuit in Fig. 7.29(d). The signal current in the base will be triangular, with a peak value $\hat{i}_b$ of

$$\hat{i}_b = \frac{\hat{v}_i}{R_{BB} + r_\pi} = \frac{0.8}{100 + 1.09} = 0.008 \text{ mA}$$

This triangular-wave current will be superimposed on the quiescent base current I_B, as shown in Fig. 7.30(b). The base–emitter voltage will consist of a triangular-wave component superimposed on the dc V_{BE} that is approximately 0.7 V. The peak value of the triangular waveform will be

$$\hat{v}_{be} = \hat{v}_i \frac{r_\pi}{r_\pi + R_{BB}} = 0.8 \frac{1.09}{100 + 1.09} = 8.6 \text{ mV}$$

The total v_{BE} is sketched in Fig. 7.30(c).

The signal current in the collector will be triangular in waveform, with a peak value $\hat{i}_c$ given by

$$\hat{i}_c = \beta \hat{i}_b = 100 \times 0.008 = 0.8 \text{ mA}$$

This current will be superimposed on the quiescent collector current I_C ($= 2.3$ mA), as shown in Fig. 7.30(d).

We can find the signal voltage at the collector by multiplying v_i by the voltage gain; that is,

$$\hat{v}_o = 3.04 \times 0.8 = 2.43 \text{ V}$$

Figure 7.30(e) shows a sketch of the total collector voltage v_C versus time. Note the phase reversal between the input signal v_i and the output signal v_o.

Finally, we observe that each of the total quantities is the sum of a dc quantity (found from the dc circuit in Fig. 7.29b), and a signal quantity (found from the circuit in Fig. 7.29d).

Example 7.6 *continued*

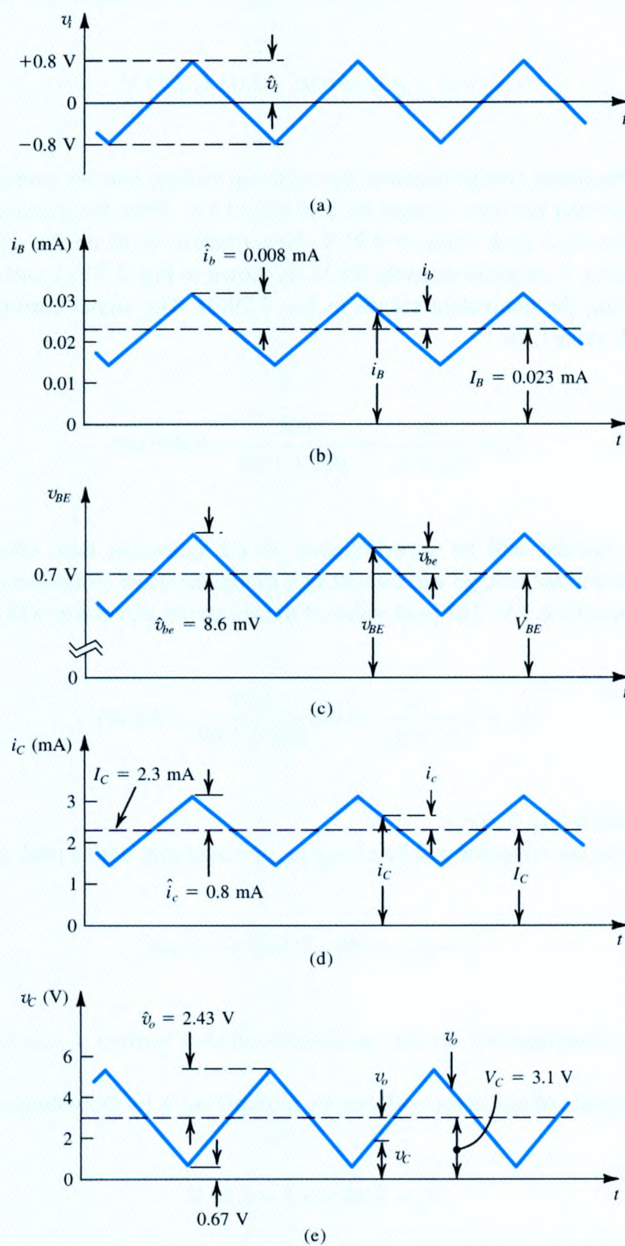


Figure 7.30 Signal waveforms in the circuit of Fig. 7.29.

Example 7.7

We need to analyze the circuit of Fig. 7.31(a) to determine the voltage gain and the signal waveforms at various points. Capacitor C_{C1} couples the signal v_i to the emitter while blocking dc. In this way the dc bias established by V^+ and V^- together with R_E and R_C will not be disturbed when the signal v_i is connected.

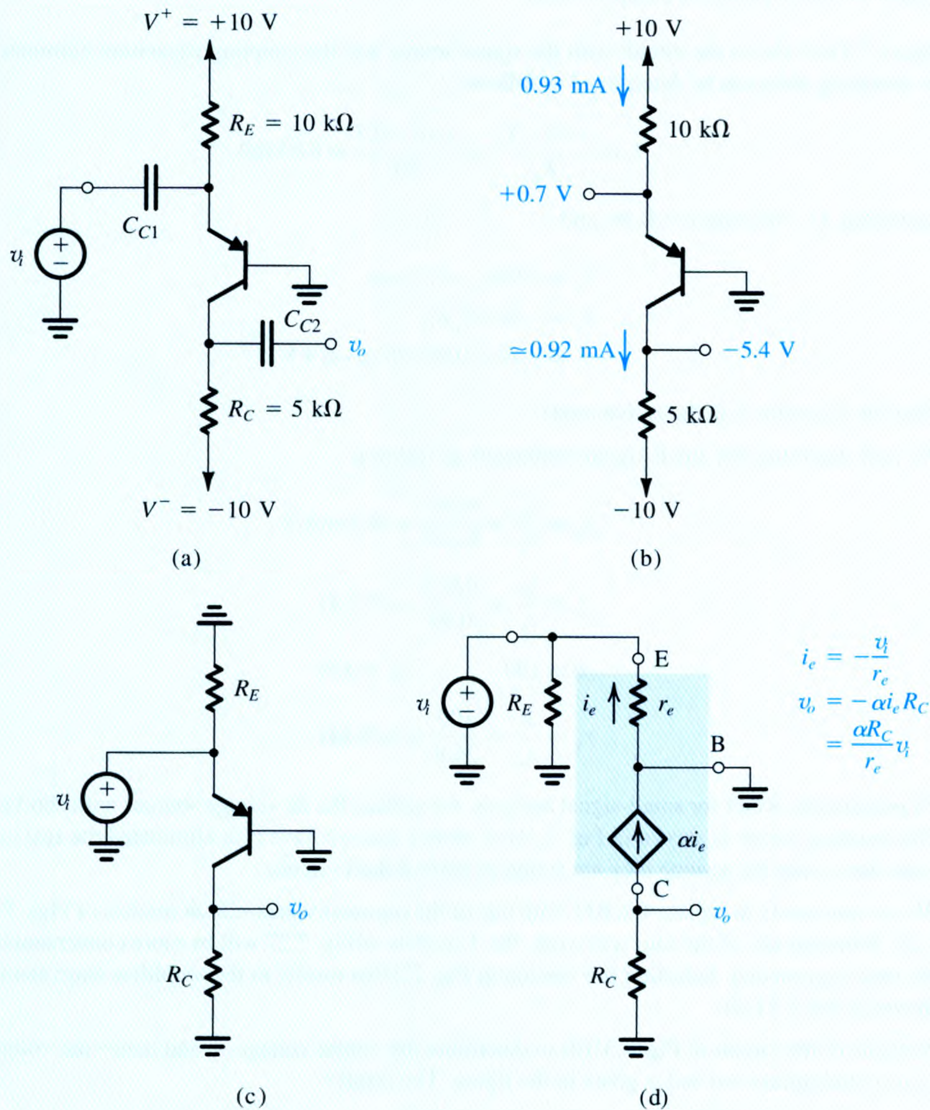


Figure 7.31 Example 7.7: (**a**) circuit; (**b**) dc analysis; (**c**) circuit with the dc sources eliminated; (**d**) small-signal analysis using the T model for the BJT.

Example 7.7 *continued*

Assume C_{C1} to be very large so as to act as a perfect short circuit at signal frequencies of interest. Similarly, another very large capacitor C_{C2} is used to couple the output signal v_o to other parts of the system. You may neglect the Early effect.

Solution

Here again we will follow a five-step process:

1. Figure 7.31(b) shows the circuit with the signal source and the coupling capacitors eliminated. The dc operating point can be determined as follows:

$$I_E = \frac{+10 - V_E}{R_E} \simeq \frac{+10 - 0.7}{10} = 0.93 \text{ mA}$$

Assuming $\beta = 100$, then $\alpha = 0.99$, and

$$I_C = 0.99 I_E = 0.92 \text{ mA}$$
$$V_C = -10 + I_C R_C$$
$$= -10 + 0.92 \times 5 = -5.4 \text{ V}$$

Thus the transistor is in the active mode.

2. We now determine the small-signal parameters as follows:

$$g_m = \frac{I_C}{V_T} = \frac{0.92}{0.025} = 36.8 \text{ mA/V}$$

$$r_e = \frac{V_T}{I_E} = \frac{0.025}{0.92} = 27.2\ \Omega$$

$$\beta = 100 \qquad \alpha = 0.99$$

$$r_\pi = \frac{\beta}{g_m} = \frac{100}{36.8} = 2.72 \text{ k}\Omega$$

3. To prepare the circuit for small-signal analysis, we replace the dc voltage sources with short circuits. The resulting circuit is shown in Fig. 7.31(c). Notice that we have also eliminated the two coupling capacitors, since we assume they are acting as perfect short circuits.

4. We are now ready to replace the BJT with one of the four equivalent-circuit models of Figs. 7.25 and 7.27. Although any of the four will work, the T models of Fig. 7.27 will be more convenient because the base is grounded. Selecting the version in Fig. 7.27(b) results in the amplifier equivalent circuit shown in Fig. 7.31(d).

5. Analysis of the circuit in Fig. 7.31(d) to determine the output voltage v_o and hence the voltage gain v_o/v_i is straightforward and is given in the figure. The result is

$$A_v = \frac{v_o}{v_i} = \frac{\alpha R_C}{r_e} = \frac{0.99 \times 5}{0.0272} = 182 \text{ V/V}$$

Note that the voltage gain is positive, indicating that the output is in phase with the input signal. This is due to the fact that the input signal is applied to the emitter rather than to the base, as was done in Example 7.5. We should emphasize that the positive gain has nothing to do with the fact that the transistor used in this example is of the *pnp* type.

Returning to the question of allowable signal magnitude, we see from Fig. 7.31(d) that $v_{eb} = v_i$. Thus, if small-signal operation is desired (for linearity), then the peak of v_i should be limited to approximately 10 mV. With $\hat{V}_i$ set to this value, as shown for a sine-wave input in Fig. 7.32, the peak amplitude at the collector, $\hat{V}_o$, will be

$$\hat{V}_o = 182 \times 0.01 = 1.82 \text{ V}$$

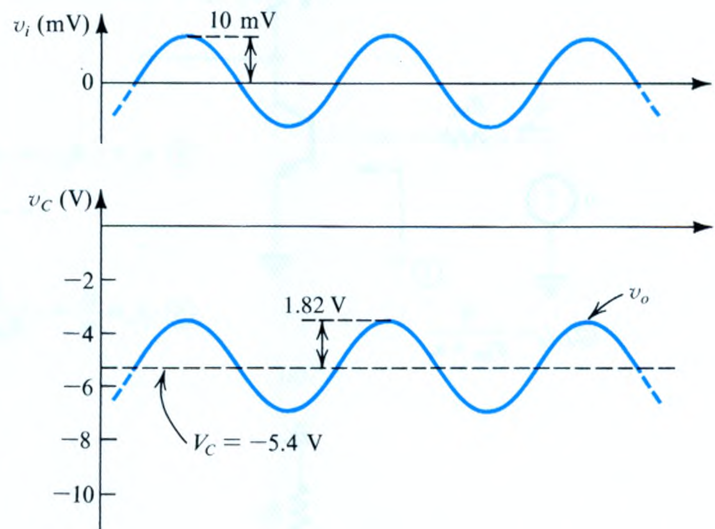


Figure 7.32 Input and output waveforms for the circuit of Fig. 7.31. Notice that this amplifier is noninverting, a property of the grounded-base configuration.

EXERCISE

7.19 To increase the voltage gain of the amplifier analyzed in Example 7.7, the collector resistance R_C is increased to 7.5 kΩ. Find the new values of V_C, A_v, and the peak amplitude of the output sine wave corresponding to an input sine wave v_i of 10-mV peak.

Ans. –3.1 V; 276 V/V; 2.76 V

Performing Small-Signal Analysis Directly on the Circuit Diagram In most cases we should explicitly replace each BJT with its small-signal model and analyze the resulting circuit, as we have done in the examples above. This systematic procedure is particularly recommended for beginning students. Experienced circuit designers, however, often perform a first-order analysis directly on the circuit. Figure 7.33 illustrates this process for the two circuits we analyzed in Examples 7.5 and 7.7. We urge you to follow this direct analysis procedure (the steps are numbered). Notice that the equivalent-circuit model is *implicitly* utilized; we are only saving the step of drawing the circuit with the BJT replaced with its

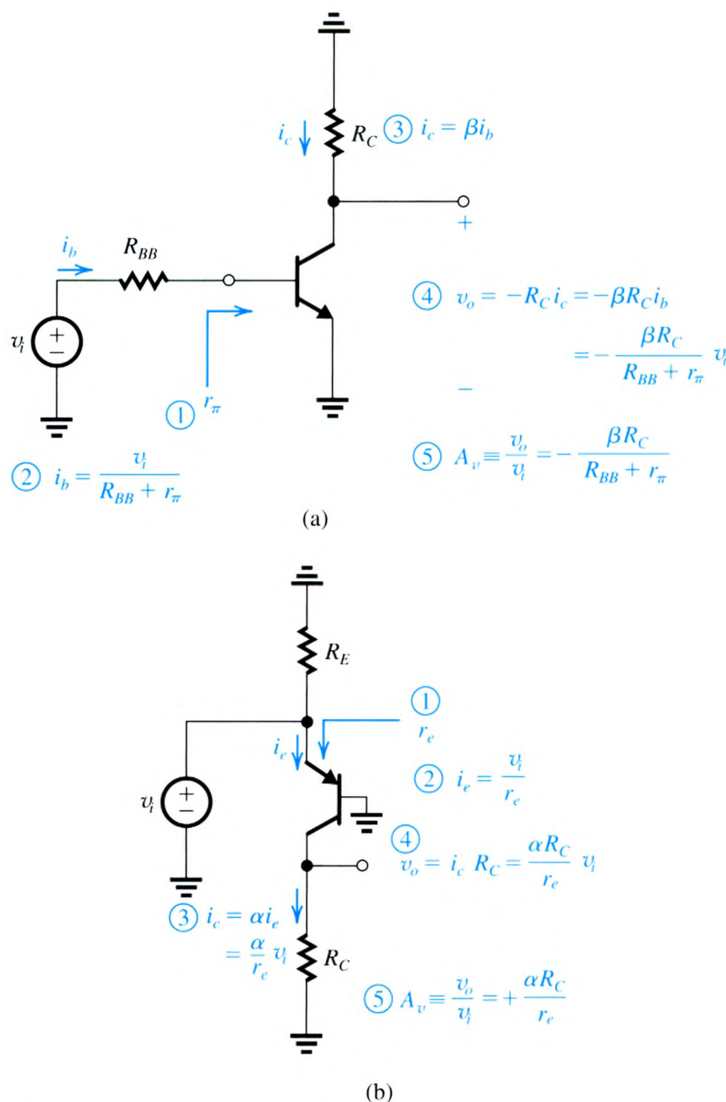


Figure 7.33 Performing signal analysis directly on the circuit diagram with the BJT small-signal model implicitly employed: (**a**) circuit for Example 7.5; (**b**) circuit for Example 7.7.

model. Direct analysis, however, has an additional very important benefit: It provides insight regarding the signal transmission through the circuit. Such insight can prove invaluable in design, particularly at the stage of selecting a circuit configuration appropriate for a given application. Direct analysis can also be used for MOS amplifier circuits.

EXERCISE

7.20 The transistor in Fig. E7.20 is biased with a constant current source $I = 1$ mA and has $\beta = 100$ and $V_A = 100$ V.

(a) Neglecting the Early effect, find the dc voltages at the base, emitter, and collector.

(b) Find g_m, r_π, and r_o.

(c) If terminal Z is connected to ground, X to a signal source v_{sig} with a source resistance R_{sig} = 2 kΩ, and Y to an 8-kΩ load resistance, use the hybrid-π model shown in Fig. 7.26 to draw the small-signal equivalent circuit of the amplifier. (Note that the current source I should be replaced with an open circuit.) Calculate the overall voltage gain v_y/v_{sig}. If r_o is neglected, what is the error in estimating the gain magnitude? (*Note:* An infinite capacitance is used to indicate that the capacitance is large enough to act as a short circuit at all signal frequencies of interest. However, the capacitor still blocks dc.)

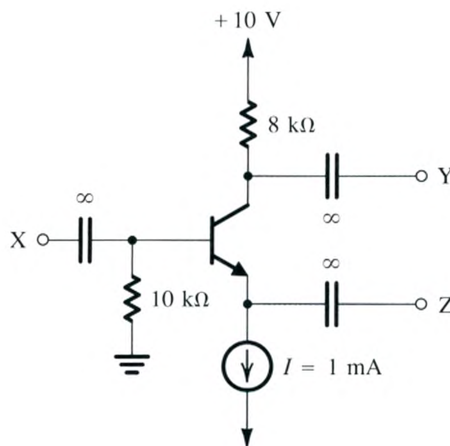


Figure E7.20

Ans. (a) –0.1 V, –0.8 V, +2.1 V; (b) 40 mA/V, 2.5 kΩ, 100 kΩ; (c) –77 V/V, +3.9%

7.2.3 Summary Tables

The results of this section are summarized in three tables included in the Summary Tables supplement found online at www.oup.com/he/sedra-smith8e. Table 7.1 lists the five steps for analyzing a MOSFET or a BJT amplifier circuit. Table 7.2 presents the MOSFET small-signal, equivalent-circuit models, together with the formulas for calculating the parameter values of the models. Finally, Table 7.3 supplies the corresponding data for the BJT.

7.3 Basic Configurations

Let's take stock of where we are and where we are going in our study of transistor amplifiers. In Section 7.1 we examined the principle underlying the application of the MOSFET, and of the BJT, as an amplifier. There we found that we can achieve almost-linear amplification by dc biasing the transistor at an appropriate point in its active region of operation, and by keeping the input signal (v_{gs} or v_{be}) small. We then developed, in Section 7.2, circuit models that represent the small-signal operation of each of the two transistor types, thus providing a systematic procedure for analyzing transistor amplifiers.

We are now ready to consider the various possible configurations of MOSFET and BJT amplifiers. In order to focus on the most important features of the various configurations, we shall present them in their most simple, or "stripped-down," version. Thus, we will not show the dc biasing arrangements, leaving the study of bias design to the next section. Section 7.5 brings everything together by showing practical *discrete-circuit amplifiers*, namely, amplifier circuits that can be constructed using discrete components. The study of integrated-circuit amplifiers begins in Chapter 8.

7.3.1 The Three Basic Configurations

There are three basic configurations for connecting a MOSFET or a BJT as an amplifier. Each of these configurations is obtained by connecting one of the device terminals to ground, thus creating a two-port network with the grounded terminal being *common* to the input and output ports. The resulting configurations are shown in Fig. 7.34(a–c) for the MOSFET and in Fig. 7.34(d–f) for the BJT.

In the circuit of Fig. 7.34(a) we have connected the source terminal to ground, applied the input voltage signal v_i between the gate and ground, and taken the output voltage signal v_o between the drain and ground, across the resistance R_D. We call this configuration the grounded-source or **common-source (CS)** amplifier, and it is by far the most popular MOS amplifier configuration; in fact, we used it in Sections 7.1 and 7.2 to study MOS amplifier operation. A parallel set of remarks apply to the BJT counterpart, the grounded-emitter or **common-emitter (CE)** amplifier in Fig. 7.34(d).

The **common-gate (CG)** or grounded-gate amplifier is shown in Fig. 7.34(b), and its BJT counterpart, the **common-base (CB)** or grounded-base amplifier in Fig. 7.34(e). Here the gate (base) is grounded, the input signal v_i is applied to the source (emitter), and the output signal v_o is taken at the drain (collector) across the resistance R_D (R_C). We encountered a CG amplifier in Example 7.4 and a CB amplifier in Example 7.7.

Finally, Fig. 7.34(c) shows the **common-drain (CD)** or grounded-drain amplifier, and Fig. 7.34(f) shows its BJT counterpart, the **common-collector (CC)** or grounded collector amplifier. Here the drain (collector) is grounded, the input signal v_i is applied between gate

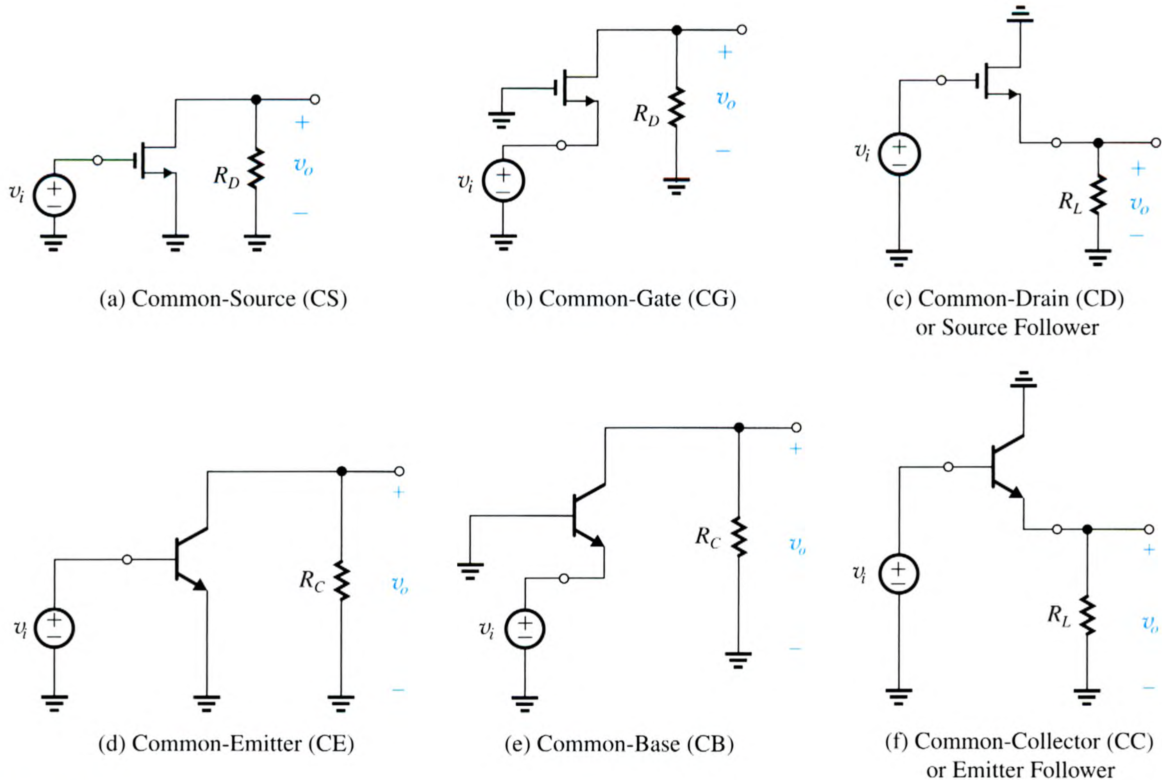


Figure 7.34 The basic configurations of transistor amplifiers: (**a**)–(**c**) for the MOSFET; (**d**)–(**f**) for the BJT.

(base) and ground, and the output voltage v_o is taken between the source (emitter) and ground, across a resistance R_L. For reasons that will become apparent shortly, this pair of configurations is more commonly called the **source follower** and the **emitter follower**.

Our study of the three basic amplifier configurations of the MOSFET and of the BJT will reveal that each has distinctly different attributes, hence areas of application. As well, we will see that although each pair of configurations (e.g., CS and CE) has many common attributes, there are important differences as well.

Our next step is to replace the transistor in each of the six circuits in Fig. 7.34 with an appropriate equivalent-circuit model and analyze the resulting circuits to determine important parameters of each amplifier configuration. To simplify matters, we will not include r_o in the initial analysis. At the end of the section we will explain when to include r_o in the analysis and how great we can expect its effect to be.

7.3.2 Characterizing Amplifiers

Before we begin our study of the different transistor amplifier configurations, let's consider how to characterize the performance of an amplifier as a circuit building block. We introduced this topic in Section 1.5.

Figure 7.35(a) shows an amplifier fed with a signal source that has an open-circuit voltage v_{sig} and an internal resistance R_{sig}. These can be the parameters of an actual signal source or, in a cascade amplifier, the Thévenin equivalent of the output circuit of another amplifier stage

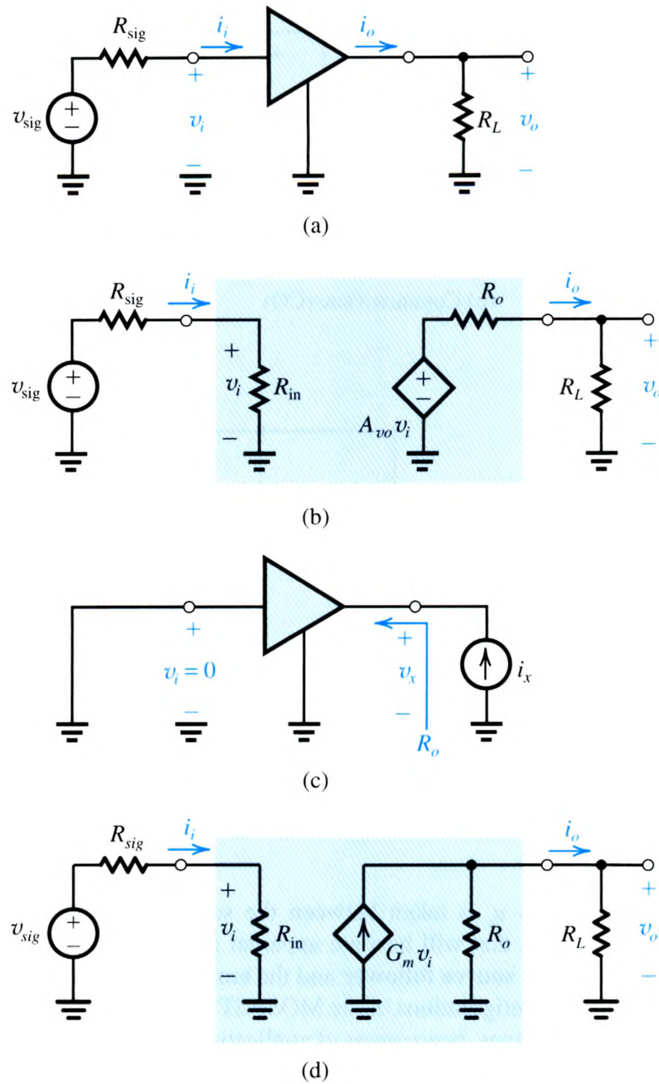


Figure 7.35 Characterization of the amplifier as a functional block: **(a)** An amplifier fed with a voltage signal v_{sig} having a source resistance R_{sig}, and feeding a load resistance R_L; **(b)** equivalent-circuit representation of the circuit in **(a)**; **(c)** determining the amplifier output resistance R_o; **(d)** alternative representation of the circuit in **(a)**.

preceding the one under study. The amplifier is shown with a load resistance R_L connected to the output terminal. Here, R_L can be an actual load resistance or the input resistance of a succeeding amplifier stage in a cascade amplifier.

Figure 7.35(b) shows the amplifier circuit with the amplifier block replaced by its equivalent-circuit model. The input resistance R_{in} represents the loading effect of the amplifier input on the signal source. It is found from

$$R_{in} \equiv \frac{v_i}{i_i}$$

and together with the resistance R_{sig} forms a voltage divider that reduces v_{sig} to the value v_i that appears at the amplifier input,

$$v_i = \frac{R_{\text{in}}}{R_{\text{in}} + R_{\text{sig}}} v_{\text{sig}} \tag{7.83}$$

Most of the amplifier circuits studied in this section are **unilateral**. That is, they do not contain internal feedback, and thus R_{in} is independent of R_L. However, in general R_{in} may depend on the load resistance R_L. Indeed one of the six configurations studied in this section, the emitter follower, exhibits such dependence.

The second parameter in characterizing amplifier performance is the **open-circuit voltage gain** A_{vo}, defined as

$$A_{vo} \equiv \left. \frac{v_o}{v_i} \right|_{R_L = \infty}$$

The third and final parameter is the output resistance R_o. Observe from Fig. 7.35(b) that R_o is the resistance seen looking back into the amplifier output terminal with v_i set to zero. Thus R_o can be determined, at least conceptually, as indicated in Fig. 7.35(c), with

$$R_o = \frac{v_x}{i_x}$$

Because R_o is determined with $v_i = 0$, the value of R_o does not depend on R_{sig}.

The controlled source $A_{vo}v_i$ and the output resistance R_o represent the Thévenin equivalent of the amplifier output circuit, and the output voltage v_o can be found from

$$v_o = \frac{R_L}{R_L + R_o} A_{vo} v_i \tag{7.84}$$

Thus the **voltage gain of the amplifier proper**, A_v, can be found as

$$A_v \equiv \frac{v_o}{v_i} = A_{vo} \frac{R_L}{R_L + R_o} \tag{7.85}$$

and the **overall voltage gain**, G_v,

$$G_v \equiv \frac{v_o}{v_{\text{sig}}}$$

can be determined by combining Eqs. (7.83) and (7.85):

$$G_v = \frac{R_{\text{in}}}{R_{\text{in}} + R_{\text{sig}}} A_{vo} \frac{R_L}{R_L + R_o} \tag{7.86}$$

In some cases, we will find it more convenient to represent the amplifier output with the equivalent circuit shown in Fig. 7.35(d). Here, G_m is the **short-circuit transconductance** of the amplifier,

$$G_m \equiv \left. \frac{i_o}{v_i} \right|_{v_o = 0}$$

while R_o is the same as in the Thévenin representation. We can use the model in Fig. 7.35(d) to determine the open-circuit voltage gain of the amplifier proper as

$$A_{vo} \equiv \left. \frac{v_o}{v_i} \right|_{R_L = \infty} = G_m R_o$$

7.3.3 The Common-Source (CS) and Common-Emitter (CE) Amplifiers

Of the three basic transistor amplifier configurations, the common-source (or common-emitter for BJT) is the most widely used. Typically, in an amplifier formed by cascading a number of gain stages, the bulk of the voltage gain is obtained by using one or more common-source (or common-emitter) stages in cascade.

Characteristic Parameters of the CS Amplifier Figure 7.36(a) shows a common-source amplifier (with the biasing arrangement omitted) fed with a signal source v_{sig} having a source resistance R_{sig}. We want to analyze this circuit to determine R_{in}, A_{vo}, and R_o. For this purpose, we assume that R_D is part of the amplifier; thus if a load resistance R_L is connected to the amplifier output, R_L appears in parallel with R_D. In such a case, we want to determine A_v and G_v as well.

Replacing the MOSFET with its hybrid-π model (without r_o), we obtain the CS amplifier equivalent circuit in Fig. 7.36(b) for which, tracing the signal from input to output, we can write by inspection

$$R_{in} = \infty \tag{7.87}$$

$$v_i = v_{sig}$$

$$v_{gs} = v_i$$

$$v_o = -g_m v_{gs} R_D$$

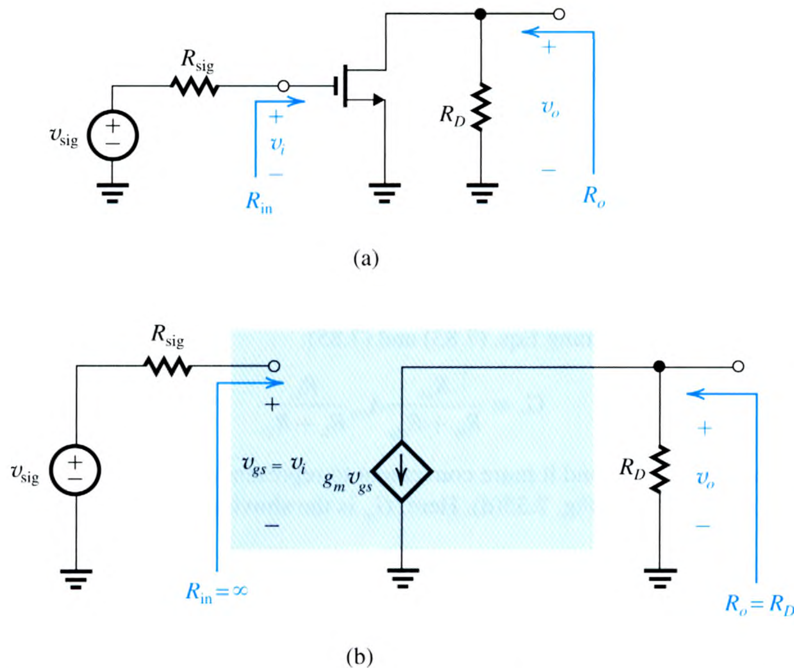


Figure 7.36 **(a)** Common-source amplifier fed with a signal v_{sig} from a generator with a resistance R_{sig}. The bias circuit is omitted. **(b)** The common-source amplifier with the MOSFET replaced with its hybrid-π model.

Thus,

$$A_{vo} \equiv \frac{v_o}{v_i} = -g_m R_D \tag{7.88}$$

$$R_o = R_D \tag{7.89}$$

If a load resistance R_L is connected across R_D, the voltage gain A_v can be obtained from

$$A_v = A_{vo} \frac{R_L}{R_L + R_o} \tag{7.90}$$

where A_{vo} is given by Eq. (7.88) and R_o by Eq. (7.89), or alternatively by simply adding R_L in parallel with R_D in Eq. (7.88), thus

$$A_v = -g_m(R_D \| R_L) \tag{7.91}$$

We can easily show that the expression obtained from Eq. (7.90) is identical to that in Eq. (7.91). Finally, since $R_{in} = \infty$ and thus $v_i = v_{sig}$, the overall voltage gain G_v is equal to A_v,

$$G_v \equiv \frac{v_o}{v_{sig}} = -g_m(R_D \| R_L) \tag{7.92}$$

EXERCISE

7.21 A CS amplifier utilizes a MOSFET biased at $I_D = 0.25$ mA with $V_{OV} = 0.25$ V and $R_D = 20$ kΩ. The amplifier is fed with a signal source having $R_{sig} = 100$ kΩ, and a 20-kΩ load is connected to the output. Find R_{in}, A_{vo}, R_o, A_v, and G_v. If, to maintain reasonable linearity, the peak of the input sine-wave signal is limited to 10% of $2V_{OV}$, what is the peak of the sine-wave voltage at the output?
Ans. ∞; −40 V/V; 20 kΩ; −20 V/V; −20 V/V; 1 V

Characteristic Parameters of the CE Amplifier Figure 7.37(a) shows a common-emitter amplifier. Its equivalent circuit, obtained by replacing the BJT with its hybrid-π model (without r_o), is shown in Fig. 7.37(b). We can analyze this circuit to obtain the characteristic parameters of the CE amplifier. The analysis parallels that for the MOSFET above except that here we have the added complexity of a finite input resistance r_π. Tracing the signal through the amplifier from input to output, we can write by inspection

$$\begin{aligned} R_{in} &= r_\pi \\ v_i &= \frac{r_\pi}{r_\pi + R_{sig}} v_{sig} \\ v_\pi &= v_i \\ v_o &= -g_m v_\pi R_C \end{aligned} \tag{7.93}$$

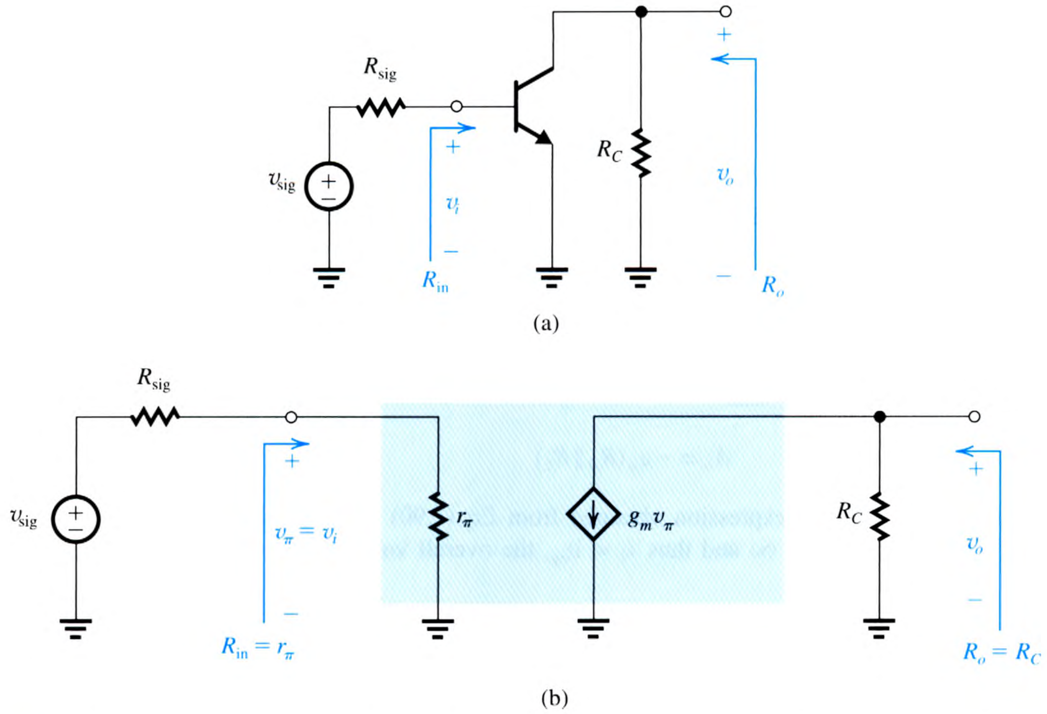


Figure 7.37 (**a**) Common-emitter amplifier fed with a signal v_{sig} from a generator with a resistance R_{sig}. The bias circuit is omitted. (**b**) The common-emitter amplifier circuit with the BJT replaced by its hybrid-π model.

Thus,

$$A_{vo} \equiv \frac{v_o}{v_i} = -g_m R_C \tag{7.94}$$

$$R_o = R_C \tag{7.95}$$

With a load resistance R_L connected across R_C,

$$A_v = -g_m(R_C \parallel R_L) \tag{7.96}$$

and the overall voltage gain G_v can be found from

$$G_v \equiv \frac{v_o}{v_{sig}} = \frac{v_i}{v_{sig}}\frac{v_o}{v_i}$$

Thus,

$$G_v = -\frac{r_\pi}{r_\pi + R_{sig}} g_m(R_C \parallel R_L) \tag{7.97}$$

It is important to note here that the finite input resistance (r_π) reduces the magnitude of the voltage gain by the voltage-divider ratio $r_\pi/(r_\pi + R_{sig})$. The extent of the gain reduction depends on the relative values of r_π and R_{sig}. However, there is a compensating effect in

the CE amplifier: g_m of the BJT is usually much higher than the corresponding value of the MOSFET.

Example 7.8

A CE amplifier using a BJT with $\beta = 100$ is biased at $I_C = 1$ mA and has a collector resistance $R_C = 5\ \text{k}\Omega$. Find R_{in}, R_o, and A_{vo}. If the amplifier is fed with a signal source having a resistance of 5 kΩ, and a load resistance $R_L = 5\ \text{k}\Omega$ is connected to the output terminal, find the resulting A_v and G_v. If $\hat{v}_\pi$ is to be limited to 5 mV, what are the corresponding $\hat{v}_{\text{sig}}$ and $\hat{v}_o$ with the load connected?

Solution

At $I_C = 1$ mA,

$$g_m = \frac{I_C}{V_T} = \frac{1\text{ mA}}{0.025\text{ V}} = 40\text{ mA/V}$$

$$r_\pi = \frac{\beta}{g_m} = \frac{100}{40\text{ mA/V}} = 2.5\ \text{k}\Omega$$

The amplifier characteristic parameters can now be found as

$$R_{\text{in}} = r_\pi = 2.5\ \text{k}\Omega$$

$$A_{vo} = -g_m R_C$$

$$= -40\text{ mA/V} \times 5\ \text{k}\Omega$$

$$= -200\text{ V/V}$$

$$R_o = R_C = 5\ \text{k}\Omega$$

With a load resistance $R_L = 5\ \text{k}\Omega$ connected at the output, we can find A_v by either of two approaches:

$$A_v = A_{vo}\frac{R_L}{R_L + R_o}$$

$$= -200 \times \frac{5}{5+5} = -100\text{ V/V}$$

or

$$A_v = -g_m(R_C \parallel R_L)$$

$$= -40(5 \parallel 5) = -100\text{ V/V}$$

The overall voltage gain G_v can now be found as

$$G_v = \frac{R_{\text{in}}}{R_{\text{in}} + R_{\text{sig}}} A_v$$

$$= \frac{2.5}{2.5+5} \times -100 = -33.3\text{ V/V}$$

Example 7.8 *continued*

If the maximum amplitude of v_π is to be 5 mV, the corresponding value of $\hat{v}_{sig}$ is

$$\hat{v}_{sig} = \left(\frac{R_{in} + R_{sig}}{R_{in}}\right)\hat{v}_\pi = \frac{2.5+5}{2.5} \times 5 = 15 \text{ mV}$$

and the amplitude of the signal at the output is

$$\hat{v}_o = G_v \hat{v}_{sig} = 33.3 \times 0.015 = 0.5 \text{ V}$$

EXERCISE

7.22 The designer of the amplifier in Example 7.8 decides to lower the bias current to half its original value in order to raise the input resistance and hence increase the fraction of v_{sig} that appears at the input of the amplifier proper. In an attempt to maintain the voltage gain, the designer decides to double the value of R_C. For the new design, determine R_{in}, A_{vo}, R_o, A_v, and G_v. If the peak amplitude of v_π is to be limited to 5 mV, what are the corresponding values of $\hat{v}_{sig}$ and $\hat{v}_o$ (with the load connected)?
Ans. 5 kΩ; −200 V/V; 10 kΩ; −66.7 V/V; −33.3 V/V; 10 mV; 0.33 V

Comment: Although a larger fraction of the input signal reaches the amplifier input, linearity considerations cause the output signal to be in fact smaller than in the original design!

Final Remarks

1. The CS and CE amplifiers are the most useful of all transistor amplifier configurations. They exhibit a moderate-to-high input resistance (infinite for the CS), a moderate-to-high output resistance, and reasonably high voltage gain.
2. The input resistance of the CE amplifier, $R_{in} = r_\pi = \beta/g_m$, is inversely proportional to the dc bias current I_C. To increase R_{in}, we are tempted to lower the bias current I_C; however, this also lowers g_m and hence the voltage gain. This is a significant design trade-off. If we want a much higher input resistance, then we can modify the CE configuration (as discussed in Section 7.3.4), or we can insert an emitter-follower stage between the signal source and the CE amplifier (see Section 7.3.6).
3. Reducing R_D or R_C to lower the output resistance of the CS or CE amplifier, respectively, is usually not a viable proposition because this also reduces the voltage gain. Alternatively, if we need a very low output resistance (in the ohms or tens-of-ohms range), we can use a source-follower or an emitter-follower stage between the output of the CS or CE amplifier and the load resistance (see Section 7.3.6).

4. Although the CS and the CE configurations are the workhorses of transistor amplifiers, both suffer from a limitation on their high-frequency response. As we will show in Chapter 10, combining the CS (CE) amplifier with a CG (CB) amplifier can extend the bandwidth considerably. The CG and CB amplifiers are studied in Section 7.3.5.

7.3.4 The Common-Source (Common-Emitter) Amplifier with a Source (Emitter) Resistance

It is often beneficial to insert a resistance R_s (a resistance R_e) in the source lead (the emitter lead) of a common-source (common-emitter) amplifier. Figure 7.38(a) shows a CS amplifier with a resistance R_s in its source lead. The corresponding small-signal equivalent circuit is shown in Fig. 7.38(b), where we have used the T model for the MOSFET. We use the T model instead of the hybrid-π model because it makes the analysis in this case much simpler. In general, *whenever a resistance is connected in the source lead, the T model is preferred.* The source resistance then appears in series with the model resistance $1/g_m$ and can be added to it.

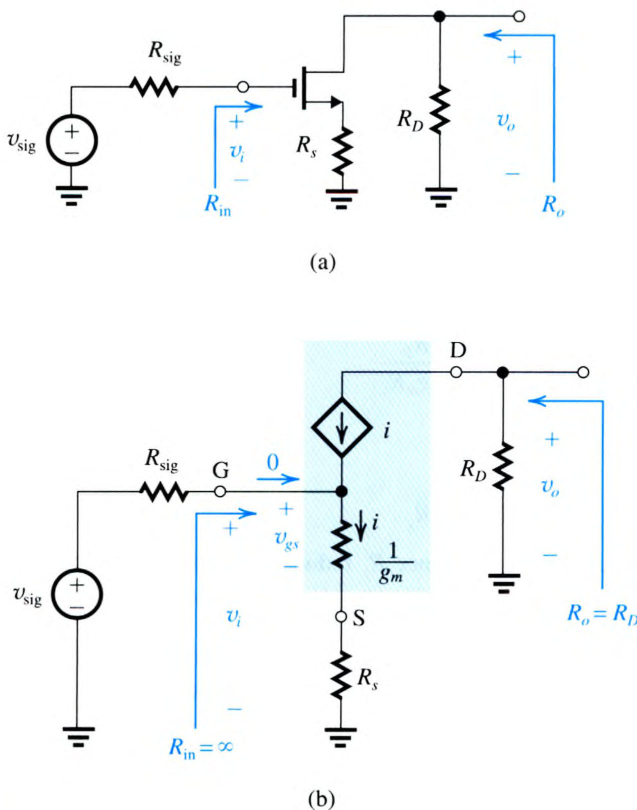


Figure 7.38 The CS amplifier with a source resistance R_s: **(a)** circuit without bias details; **(b)** equivalent circuit with the MOSFET represented by its T model.

From Fig. 7.38(b) we see that as expected, the input resistance R_{in} is infinite and thus $v_i = v_{sig}$. Unlike the CS amplifier, however, here only a fraction of v_i appears between gate and source as v_{gs}. The voltage divider composed of $1/g_m$ and R_s can be used to determine v_{gs}, as follows:

$$v_{gs} = v_i \frac{1/g_m}{1/g_m + R_s} = \frac{v_i}{1 + g_m R_s} \tag{7.98}$$

Thus we can use the value of R_s to control the magnitude of the signal v_{gs} and thereby ensure that v_{gs} does not become too large and cause unacceptably high nonlinear distortion. This is the first benefit of including resistor R_s. We will see other benefits in later sections and chapters. For instance, we will show in Chapter 10 that R_s causes the useful bandwidth of the amplifier to be extended. The mechanism by which R_s causes such improvements in amplifier performance is *negative feedback*. To see how R_s introduces negative feedback, refer to Fig. 7.38(a): If with v_{sig} and hence v_i kept constant, the drain current increases for some reason, the source current also will increase, resulting in an increased voltage drop across R_s. Thus the source voltage rises, and the gate-to-source voltage decreases. The latter effect causes the drain current to decrease, counteracting the initially assumed change, an indication of the presence of negative feedback, which we will formally study in Chapter 11. There we will learn that the improvements negative feedback provides come at the expense of a reduction in gain. We will now show this to be the case in the circuit of Fig. 7.38.

We can find the output voltage v_o by multiplying the controlled-source current i by R_D,

$$v_o = -iR_D$$

We can then find the current i in the source lead by dividing v_i by the total resistance in the source,

$$i = \frac{v_i}{1/g_m + R_s} = \left(\frac{g_m}{1 + g_m R_s}\right) v_i \tag{7.99}$$

Thus, the voltage gain A_{vo} is

$$A_{vo} \equiv \frac{v_o}{v_i} = -\frac{g_m R_D}{1 + g_m R_s} \tag{7.100}$$

which can also be expressed as

$$A_{vo} = -\frac{R_D}{1/g_m + R_s} \tag{7.101}$$

Equation (7.100) indicates that including the resistance R_s reduces the voltage gain by the factor $(1 + g_m R_s)$. This is the price we pay for the improvements gained as a result of R_s. It is interesting to note that in Chapter 11, we will find that the factor $(1 + g_m R_s)$ is the "amount of negative feedback" introduced by R_s. It is also the same factor by which linearity, bandwidth, and other performance parameters improve. Because of the negative-feedback action of R_s it is known as a **source-degeneration resistance**.

There is another useful interpretation of the expression for the drain current in Eq. (7.99): The quantity between brackets on the right-hand side can be thought of as the "effective transconductance with R_s included." Thus, including R_s reduces the transconductance by the factor $(1 + g_m R_s)$. This, of course, is simply the result of the fact that only a fraction, $1/(1 + g_m R_s)$, of v_i appears as v_{gs} (see Eq. 7.98).

The alternative gain expression in Eq. (7.101) has a powerful and insightful interpretation: The voltage gain between gate and drain is equal to the ratio of the total resistance in the drain (R_D) to the total resistance in the source ($1/g_m + R_s$),

$$\text{Voltage gain from gate to drain} = -\frac{\text{Total resistance in drain}}{\text{Total resistance in source}} \tag{7.102}$$

This is a general expression. For instance, setting $R_s = 0$ in Eq. (7.101) yields A_{vo} of the CS amplifier.

Finally, we consider the situation when a load resistance R_L is connected at the output. We can find the gain A_v using the open-circuit voltage gain A_{vo} together with the output resistance R_o, which we find to be

$$R_o = R_D$$

Alternatively, we can find A_v by simply replacing R_D in Eq. (7.101) or (7.100) by ($R_D \| R_L$); thus,

$$A_v = -\frac{g_m(R_D \| R_L)}{1 + g_m R_s} \tag{7.103}$$

or

$$A_v = -\frac{R_D \| R_L}{1/g_m + R_s} \tag{7.104}$$

Observe that Eq. (7.104) is a direct application of the ratio of total resistance rule of Eq. (7.102). Note, too, that because R_{in} is infinite, $v_i = v_{sig}$ and the overall voltage gain G_v is equal to A_v.

EXERCISE

7.23 In Exercise 7.21 we applied an input signal v_{sig} of 50 mV peak and obtained an output signal of approximately 1 V peak. Assume that for some reason we now have an input signal v_{sig} that is 0.2 V peak and that we want to modify the circuit to keep v_{gs} unchanged, and thus keep the nonlinear distortion from increasing. What value should we use for R_s? What value of G_v will result? What will the peak signal at the output become? Assume $r_o = \infty$.
Ans. 1.5 kΩ; −5 V/V; 1 V

We next turn our attention to the BJT case. Figure 7.39(a) shows a CE amplifier with a resistance R_e in its emitter. The corresponding equivalent circuit, using the T model, is shown in Fig. 7.39(b). Note that in the BJT case, as with the MOSFET, the T model results in a simpler analysis and should be used whenever there is a resistance in series with the emitter.

To determine the amplifier input resistance R_{in}, we note from Fig. 7.39(b) that

$$R_{in} \equiv \frac{v_i}{i_b}$$

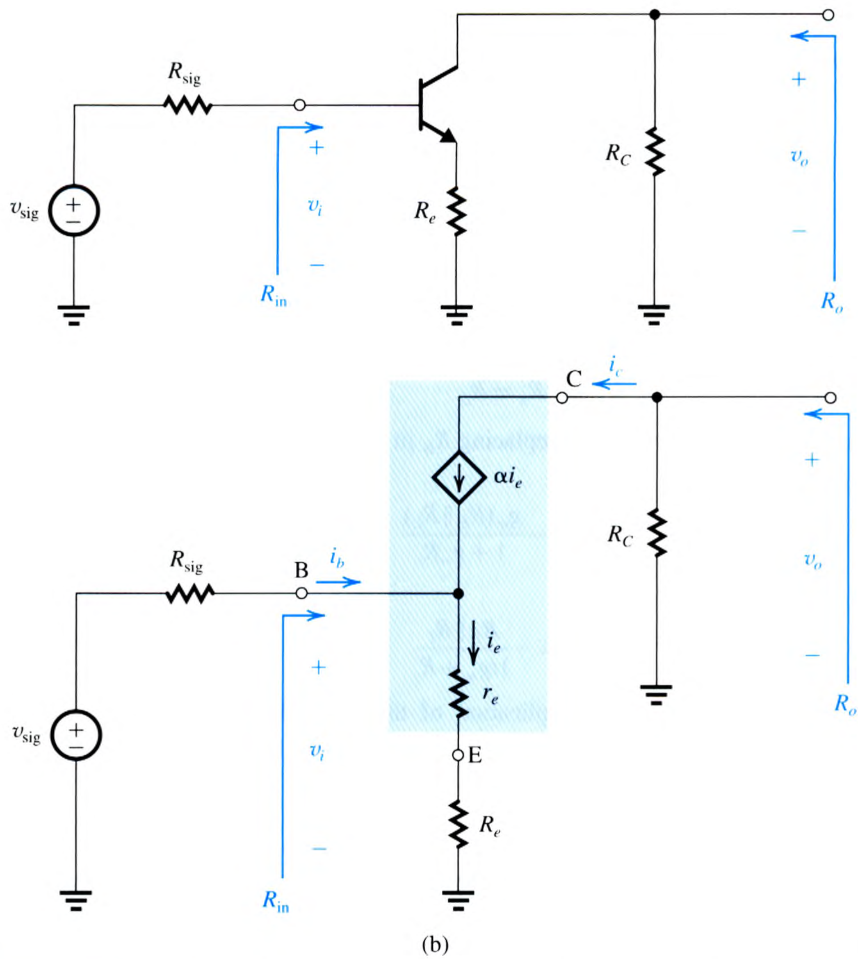


Figure 7.39 The CE amplifier with an emitter resistance R_e; (**a**) circuit without bias details; (**b**) equivalent circuit with the BJT replaced with its T model.

where

$$i_b = (1-\alpha)i_e = \frac{i_e}{\beta+1} \tag{7.105}$$

and

$$i_e = \frac{v_i}{r_e + R_e} \tag{7.106}$$

Thus,

$$R_{in} = (\beta+1)(r_e + R_e) \tag{7.107}$$

This is a very important result. It states that *the input resistance looking into the base is* $(\beta+1)$ *times the total resistance in the emitter*; this is known as the **resistance-reflection rule**. The factor $(\beta+1)$ arises because the base current is $1/(\beta+1)$ times the emitter current. The expression for R_{in} in Eq. (7.107) shows clearly that including a resistance R_e in the emitter

can substantially increase R_{in}, a very desirable result. Indeed, the value of R_{in} is increased by the ratio

$$\frac{R_{in}(\text{with } R_e \text{ included})}{R_{in}(\text{without } R_e)} = \frac{(\beta+1)(r_e+R_e)}{(\beta+1)r_e}$$
$$= 1 + \frac{R_e}{r_e} \simeq 1 + g_m R_e \tag{7.108}$$

Thus the circuit designer can use the value of R_e to control the value of R_{in}.

To determine the voltage gain A_{vo}, we see from Fig. 7.38(b) that

$$v_o = -i_c R_C$$
$$= -\alpha i_e R_C$$

Substituting for i_e from Eq. (7.106) gives

$$A_{vo} = -\alpha \frac{R_C}{r_e + R_e} \tag{7.109}$$

This is a very useful result: It states that the gain from base to collector is α times the ratio of the total resistance in the collector to the total resistance in the emitter (in this case, $r_e + R_e$),

$$\text{Voltage gain from base to collector} = -\alpha \frac{\text{Total resistance in collector}}{\text{Total resistance in emitter}} \tag{7.110}$$

This is the BJT version of the MOSFET expression in Eq. (7.102) except that here we have the additional factor α. This factor arises because $i_c = \alpha i_e$, unlike the MOSFET case where $i_d = i_s$. Usually, $\alpha \simeq 1$ and can be dropped from Eq. (7.110).

The open-circuit voltage gain in Eq. (7.109) can be expressed alternatively as

$$A_{vo} = -\frac{\alpha}{r_e} \frac{R_C}{1 + R_e/r_e}$$

Thus,

$$A_{vo} = -\frac{g_m R_C}{1 + R_e/r_e} \simeq -\frac{g_m R_C}{1 + g_m R_e} \tag{7.111}$$

Thus, including R_e reduces the magnitude of the voltage gain by the factor $(1 + g_m R_e)$, which is the same factor by which R_{in} is increased. This highlights an interesting trade-off between gain and input resistance, a trade-off that the designer can exercise through the choice of an appropriate value for R_e.

The output resistance R_o can be found from the circuit in Fig. 7.39(b) by inspection:

$$R_o = R_C$$

If a load resistance R_L is connected at the amplifier output, A_v can be found as

$$A_v = A_{vo} \frac{R_L}{R_L + R_o}$$
$$= -\alpha \frac{R_C}{r_e + R_e} \frac{R_L}{R_L + R_C}$$
$$= -\alpha \frac{R_C \| R_L}{r_e + R_e} \tag{7.112}$$

which could have been written directly using Eq. (7.110). The overall voltage gain G_v can now be found:

$$G_v = \frac{R_{\text{in}}}{R_{\text{in}} + R_{\text{sig}}} \times -\alpha \frac{R_C \| R_L}{r_e + R_e}$$

Substituting for R_{in} from Eq. (7.107) and replacing α with $\beta/(\beta+1)$ results in

$$G_v = -\beta \frac{R_C \| R_L}{R_{\text{sig}} + (\beta+1)(r_e + R_e)} \tag{7.113}$$

Careful examination of this expression reveals that the denominator comprises the total resistance in the base circuit [recall that $(\beta+1)(r_e + R_e)$ is the reflection of $(r_e + R_e)$ from the emitter side to the base side]. Thus the expression in Eq. (7.113) states that the voltage gain from base to collector is equal to β times the ratio of the total resistance in the collector to the total resistance in the base. The factor β appears because it is the ratio of the collector current to the base current. This general and useful expression has no counterpart in the MOS case. We notice that the overall voltage gain G_v is lower in magnitude than the value without R_e, namely,

$$G_v = -\beta \frac{R_C \| R_L}{R_{\text{sig}} + (\beta+1)r_e} \tag{7.114}$$

because of the additional term $(\beta+1)R_e$ in the denominator. The gain, however, is now less sensitive to the value of β, a desirable result because of the typical wide variability in the value of β.

Another important consequence of including R_e in the emitter is that it enables the amplifier to handle larger input signals without incurring nonlinear distortion. This is because only a fraction of the input signal at the base, v_i, appears between the base and the emitter. Specifically, from the circuit in Fig. 7.39(b), we see that

$$\frac{v_\pi}{v_i} = \frac{r_e}{r_e + R_e} \simeq \frac{1}{1 + g_m R_e} \tag{7.115}$$

Thus, for the same v_π, the signal at the input terminal of the amplifier, v_i, can be greater than for the CE amplifier by the factor $(1 + g_m R_e)$.

To summarize, including a resistance R_e in the emitter of the CE amplifier results in the following:

1. The input resistance R_{in} is increased by the factor $(1 + g_m R_e)$.
2. The voltage gain from base to collector, A_v, is reduced by the factor $(1 + g_m R_e)$.
3. For the same nonlinear distortion, the input signal v_i can be increased by the factor $(1 + g_m R_e)$.
4. The overall voltage gain is less dependent on the value of β.
5. The high-frequency response is significantly improved (as we will see in Chapter 10).

With the exception of gain reduction, these characteristics represent performance improvements. Indeed, the reduction in gain is the price we pay for the other performance improvements. In many cases this is a good bargain; it is the underlying philosophy for the use of negative feedback. That the resistance R_e introduces negative feedback in the amplifier circuit can be verified through a procedure similar to the one we used above for the

MOSFET case. In Chapter 11, where we will study negative feedback formally, we will find that the factor $(1+g_m R_e)$, which appears repeatedly, is the "amount of negative feedback" introduced by R_e. Finally, we note that the negative-feedback action of R_e gives it the name **emitter degeneration resistance**.

Example 7.9

For the CE amplifier specified in Example 7.8, what value of R_e is needed to raise R_{in} to a value four times that of R_{sig}? With R_e included, find A_{vo}, R_o, A_v, and G_v. Also, if $\hat{v}_\pi$ is limited to 5 mV, what are the corresponding values of $\hat{v}_{\text{sig}}$ and $\hat{v}_o$?

Solution

To obtain $R_{\text{in}} = 4R_{\text{sig}} = 4 \times 5 = 20\ \text{k}\Omega$, the required R_e is found from

$$20 = (\beta+1)(r_e + R_e)$$

With $\beta = 100$,

$$r_e + R_e \simeq 200\ \Omega$$

Thus,

$$R_e = 200 - 25 = 175\ \Omega$$

$$A_{vo} = -\alpha \frac{R_C}{r_e + R_e}$$

$$\simeq -\frac{5000}{25+175} = -25\ \text{V/V}$$

$$R_o = R_C = 5\ \text{k}\Omega \text{ (unchanged)}$$

$$A_v = A_{vo}\frac{R_L}{R_L + R_o} = -25 \times \frac{5}{5+5} = -12.5\ \text{V/V}$$

$$G_v = \frac{R_{\text{in}}}{R_{\text{in}} + R_{\text{sig}}} A_v = -\frac{20}{20+5} \times 12.5 = -10\ \text{V/V}$$

For $\hat{v}_\pi = 5$ mV,

$$\hat{v}_i = \hat{v}_\pi \left(\frac{r_e + R_e}{r_e}\right)$$

$$= 5\left(1 + \frac{175}{25}\right) = 40\ \text{mV}$$

$$\hat{v}_{\text{sig}} = \hat{v}_i \frac{R_{\text{in}} + R_{\text{sig}}}{R_{\text{in}}}$$

$$= 40\left(1 + \frac{5}{20}\right) = 50\ \text{mV}$$

$$\hat{v}_o = \hat{v}_{\text{sig}} \times |G_v|$$

$$= 50 \times 10 = 500\ \text{mV} = 0.5\ \text{V}$$

Example 7.9 *continued*

Thus, while $|G_v|$ has decreased to about a third of its original value, the amplifier is able to produce as large an output signal as before for the same nonlinear distortion.

EXERCISE

7.24 Show that with R_e included, and v_π limited to a maximum value $\hat{v}_\pi$, the maximum allowable input signal, $\hat{v}_{sig}$, is given by

$$\hat{v}_{sig} = \hat{v}_\pi \left(1 + \frac{R_e}{r_e} + \frac{R_{sig}}{r_\pi}\right)$$

If the transistor is biased at $I_C = 0.5$ mA and has a β of 100, what value of R_e is needed to permit an input signal $\hat{v}_{sig}$ of 100 mV from a source with a resistance $R_{sig} = 10$ kΩ while limiting $\hat{v}_\pi$ to 10 mV? What is R_{in} for this amplifier? If the total resistance in the collector is 10 kΩ, what G_v value results?
Ans. 350 Ω; 40.4 kΩ; −19.8 V/V

7.3.5 The Common-Gate (CG) and the Common-Base (CB) Amplifiers

Figure 7.40(a) shows a common-gate amplifier with the biasing circuit omitted. The amplifier is fed with a signal source characterized by v_{sig} and R_{sig}. Since R_{sig} appears in series with the source, it is more convenient to represent the transistor with the T model than with the π model. Doing this, we obtain the amplifier equivalent circuit shown in Fig. 7.40(b).

From inspection of the equivalent circuit of Fig. 7.40(b), we see that the input resistance

$$R_{in} = \frac{1}{g_m} \tag{7.116}$$

We should have expected this, since we are looking into the source and the gate is grounded. Typically $1/g_m$ is a few hundred ohms; thus the CG amplifier has a low input resistance.

To determine the voltage gain A_{vo}, we write at the drain node

$$v_o = -iR_D$$

and substitute for the source current i from

$$i = -\frac{v_i}{1/g_m}$$

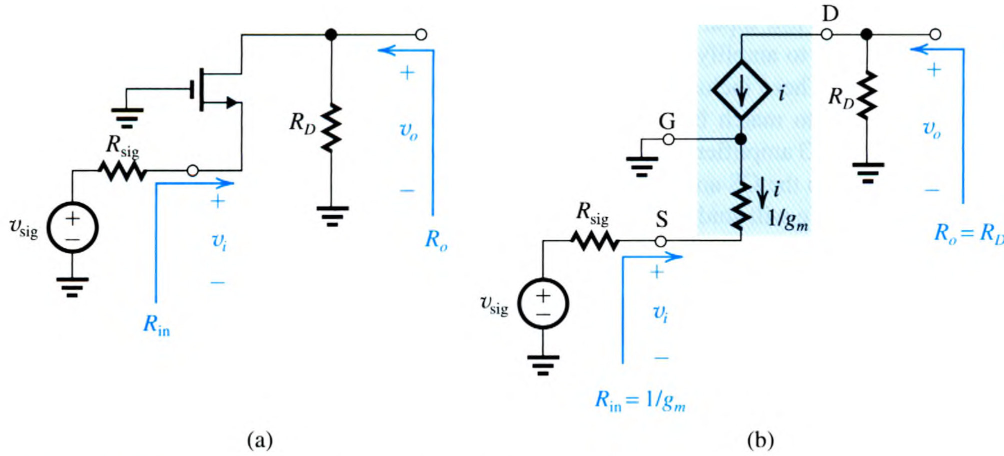


Figure 7.40 **(a)** Common-gate (CG) amplifier with bias arrangement omitted. **(b)** Equivalent circuit of the CG amplifier with the MOSFET replaced with its T model.

to obtain

$$A_{vo} \equiv \frac{v_o}{v_i} = g_m R_D \tag{7.117}$$

which except for the positive sign is identical to the expression for A_{vo} of the CS amplifier.

By inspecting the circuit in Fig. 7.40(b), we can find the output resistance of the CG circuit as

$$R_o = R_D \tag{7.118}$$

which is the same as in the case of the CS amplifier.

Although the gain of the CG amplifier proper has the same magnitude as that of the CS amplifier, this is usually not the case for the overall voltage gain. The low input resistance of the CG amplifier can cause the input signal to be severely attenuated. Specifically,

$$\frac{v_i}{v_{sig}} = \frac{R_{in}}{R_{in} + R_{sig}} = \frac{1/g_m}{1/g_m + R_{sig}} \tag{7.119}$$

from which we see that except for situations in which R_{sig} is on the order of $1/g_m$, the signal transmission factor v_i/v_{sig} can be very small and the overall voltage gain G_v can be just as small. Specifically, with a resistance R_L connected at the output

$$G_v = \frac{1/g_m}{R_{sig} + 1/g_m}\left[g_m(R_D \| R_L)\right]$$

Thus,

$$G_v = \frac{R_D \| R_L}{R_{sig} + 1/g_m} \tag{7.120}$$

Notice that *the overall voltage gain is simply the ratio of the total resistance in the drain circuit to the total resistance in the source circuit.* If R_{sig} is of the same order as R_D and R_L, G_v will be very small.

Because of its low input resistance, the CG amplifier alone has very limited application. One of its uses is to amplify high-frequency signals that come from sources with relatively low resistances. These include cables, where it is usually necessary for the input resistance of the amplifier to match the characteristic resistance of the cable. As we will show in Chapter 10, the CG amplifier has excellent high-frequency response. Thus it can be combined very usefully with the CS amplifier to take advantage of the best features of each of the two configurations. We will study a very important circuit of this kind in Chapter 8.

EXERCISE

7.25 A CG amplifier is required to match a signal source with $R_{sig} = 100\ \Omega$. At what current I_D should the MOSFET be biased if it is operated at an overdrive voltage of 0.20 V? If the total resistance in the drain circuit is 2 kΩ, what is the overall voltage gain?
Ans. 1 mA; 10 V/V

Very similar results can be obtained for the CB amplifier shown in Fig. 7.41(a). Specifically, from the equivalent circuit in Fig. 7.41(b) we can find

$$R_{in} = r_e = \frac{\alpha}{g_m} \simeq 1/g_m \tag{7.121}$$

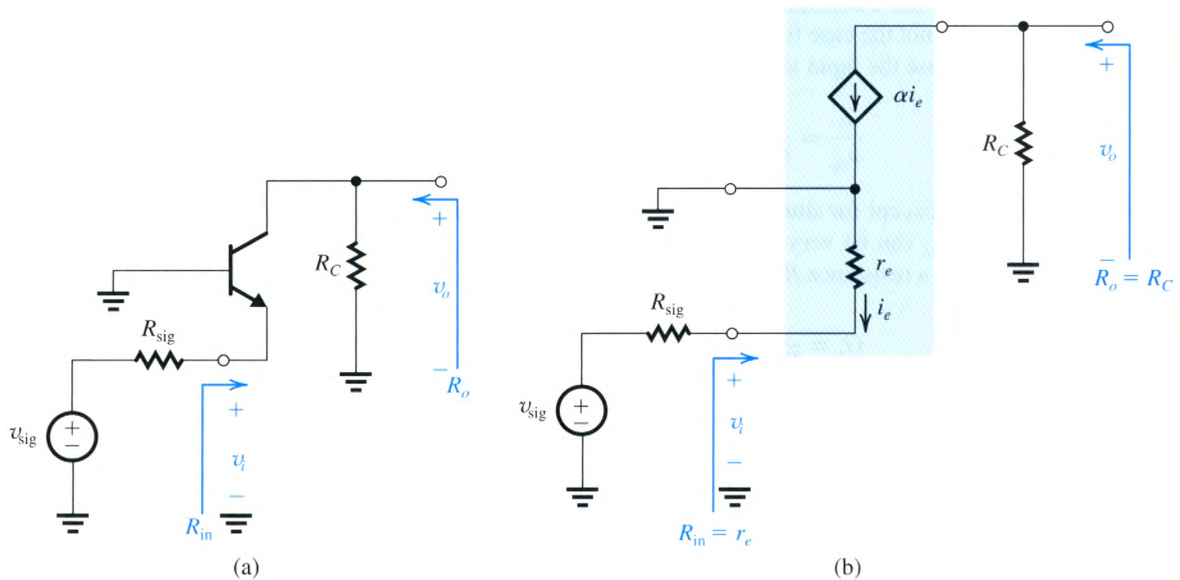


Figure 7.41 (a) CB amplifier with bias details omitted; (b) amplifier equivalent circuit with the BJT represented by its T model.

$$A_{vo} = \frac{\alpha}{r_e} R_C = g_m R_C \qquad (7.122)$$

$$R_o = R_C \qquad (7.123)$$

and with a load resistance R_L connected to the output, the overall voltage gain is

$$G_v \equiv \frac{v_o}{v_{sig}} = \alpha \frac{R_C \| R_L}{R_{sig} + r_e} \qquad (7.124)$$

Since $\alpha \simeq 1$, we see that as in the case of the CG amplifier, the overall voltage gain is simply the ratio of the total resistance in the collector to the total resistance in the emitter. We also note that the overall voltage gain is almost independent of the value of β (except through the small dependence of α on β), a desirable property. Observe that for R_{sig} of the same order as R_C and R_L, the gain will be very small.

In summary, the CB and CG amplifiers have a very low input resistance ($1/g_m$), an open-circuit voltage gain that is positive and equal in magnitude to that of the CE (CG) amplifier ($g_m R_C$ or $g_m R_D$), and, like the CE (CS) amplifier, a relatively high output resistance (R_C or R_D). Because of its very low input resistance, the CB (CG) circuit *alone* is not attractive as a voltage amplifier except in specialized applications, such as the cable amplifier mentioned above. The CB (CG) amplifier has excellent high-frequency performance, which makes it useful in combination with other circuits in the implementation of high-frequency amplifiers, as we will see in Chapters 8 and 10.

EXERCISES

7.26 Consider a CB amplifier using a BJT biased at $I_C = 1$ mA and with $R_C = 5$ kΩ. Determine R_{in}, A_{vo}, and R_o. If the amplifier is loaded in $R_L = 5$ kΩ, what value of A_v results? What G_v is obtained if $R_{sig} = 5$ kΩ?
Ans. 25 Ω; 200 V/V; 5 kΩ; 100 V/V; 0.5 V/V

7.27 A CB amplifier is required to amplify a signal delivered by a coaxial cable having a characteristic resistance of 50 Ω. What bias current I_C should be used to obtain R_{in} that is matched to the cable resistance? To get an overall voltage gain of G_v of 40 V/V, what should the total resistance in the collector (i.e., $R_C \| R_L$) be?
Ans. 0.5 mA; 4 kΩ

7.3.6 The Source and Emitter Followers

The last of the basic transistor amplifier configurations is the common-drain (common-collector) amplifier, an important circuit that finds application in the design of both small-signal amplifiers and amplifiers that are required to handle large signals and deliver substantial amounts of signal power to a load. This latter variety will be studied in Chapter 12. The common-drain amplifier is more commonly known as the *source follower*,

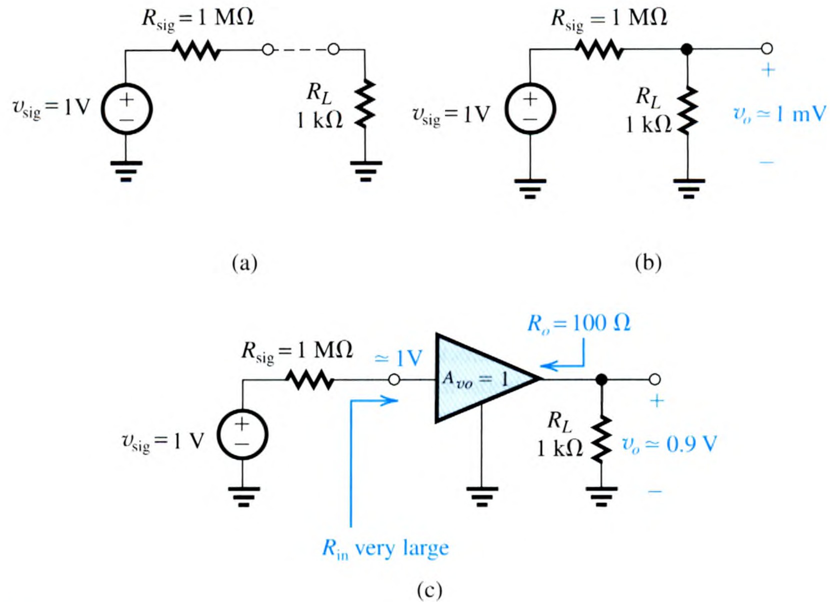


Figure 7.42 Illustrating the need for a unity-gain voltage buffer amplifier.

and the common-collector amplifier is more commonly known as the *emitter follower*. The reason behind these names will become apparent shortly.

The Need for Voltage Buffers Before analyzing the source and the emitter followers, let's look at one of their more common applications. Consider the situation depicted in Fig. 7.42(a). A signal source delivering a signal of reasonable strength (1 V) with an internal resistance of 1 MΩ is to be connected to a 1-kΩ load resistance. Connecting the source to the load directly as in Fig. 7.42(b) would result in severe attenuation of the signal; the signal appearing across the load will be only $1/(1000 + 1)$ of the input signal, or about 1 mV. An alternative course of action is shown in Fig. 7.42(c). Here we have interposed an amplifier between the source and the load. Our amplifier, however, is unlike the amplifiers we have been studying in this chapter so far; it has a voltage gain of only one. This is because our signal is already strong enough that we do not need to increase its amplitude. Note, however, that our amplifier has a very high input resistance, thus almost all of v_{sig} (i.e., 1 V) will appear at the input of the amplifier proper. Since the amplifier has a low output resistance (100 Ω), 90% of this signal (0.9 V) will appear across the load, obviously a very significant improvement over the situation without the amplifier. As we will see next, the source follower can easily implement the unity-gain buffer amplifier shown in Fig. 7.42(c).

Characteristic Parameters of the Source Follower Figure 7.43(a) shows a source follower with the bias circuit omitted. The source follower is fed with a signal generator (v_{sig}, R_{sig}) and has a load resistance R_L connected between the source terminal and ground. We shall assume that R_L includes both the actual load and any other resistance that may be present between the source terminal and ground (e.g., for biasing purposes). Normally, the actual load resistance would be much lower than these other resistances and thus would

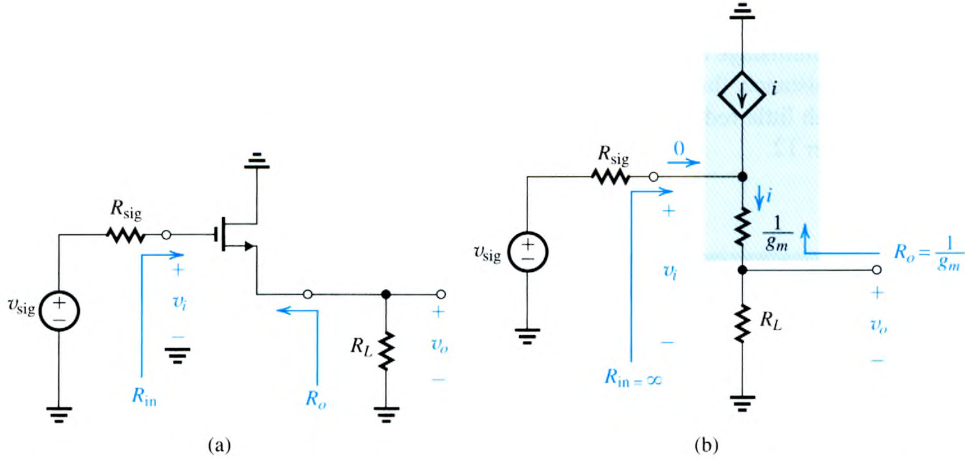


Figure 7.43 **(a)** Common-drain amplifier or source follower with the bias circuit omitted. **(b)** Equivalent circuit of the source follower obtained by replacing the MOSFET with its T model.

dominate. Since the MOSFET has a resistance R_L connected in its source terminal, it is most convenient to use the T model, as shown in Fig. 7.43(b). From this circuit we can write by inspection

$$R_{in} = \infty$$

and obtain A_v from the voltage divider formed by $1/g_m$ and R_L as

$$A_v \equiv \frac{v_o}{v_i} = \frac{R_L}{R_L + 1/g_m} \tag{7.125}$$

Setting $R_L = \infty$ we obtain

$$A_{vo} = 1 \tag{7.126}$$

We find the output resistance R_o by setting $v_i = 0$ (i.e., by grounding the gate). Now looking back into the output terminal, excluding R_L, we simply see $1/g_m$, thus

$$R_o = 1/g_m \tag{7.127}$$

The unity open-circuit voltage gain together with R_o in Eq. (7.127) can be used to find A_v when a load resistance R_L is connected. The result is simply the expression in Eq. (7.125). Finally, because of the infinite R_{in}, $v_i = v_{sig}$, and the overall voltage gain is

$$G_v = A_v = \frac{R_L}{R_L + 1/g_m} \tag{7.128}$$

Thus G_v will be lower than one. However, because $1/g_m$ is usually low, the voltage gain can be close to one. The unity open-circuit voltage gain in Eq. (7.126) indicates that the voltage at the source terminal will follow that at the input, hence the name *source follower*.

To sum up, the source follower features a very high input resistance (ideally infinite), a relatively low output resistance ($1/g_m$), and an open-circuit voltage gain that is near unity (ideally exactly one). Thus the source follower is ideally suited for implementing the

unity-gain voltage buffer of Fig. 7.42(c). The source follower is also used as the output (i.e., last) stage in a multistage amplifier, where its function is to equip the overall amplifier with a low output resistance, thus enabling it to supply relatively large load currents without loss of gain (i.e., with little reduction of output signal level). We will study the design of output stages in Chapter 12.

EXERCISES

D7.28 It is required to design a source follower that implements the buffer amplifier shown in Fig. 7.42(c). If the MOSFET is operated with an overdrive voltage $V_{OV} = 0.25$ V, at what drain current should it be biased? Find the output signal amplitude and the signal amplitude between gate and source.
Ans. 1.25 mA; 0.91 V; 91 mV

D7.29 A MOSFET is connected in the source-follower configuration and used as the output stage of a cascade amplifier. It is required to provide an output resistance of 200 Ω. If the MOSFET has $k_n' = 0.4$ mA/V^2 and is operated at $V_{OV} = 0.25$ V, find the required W/L ratio. Also specify the dc bias current I_D. If the amplifier load resistance varies over the range 1 kΩ to 10 kΩ, what is the range of G_v of the source follower?
Ans. 50; 0.625 mA; 0.83 V/V to 0.98 V/V

Characteristic Parameters of the Emitter Follower Although the emitter follower does not have an infinite input resistance (as the source follower does), it is still widely used as a voltage buffer. In fact, it is a very versatile and popular circuit. We will therefore study it in some detail.

Figure 7.44(a) shows an emitter follower with the equivalent circuit shown in Fig. 7.44(b). The input resistance R_{in} is found from

$$R_{\text{in}} = \frac{v_i}{i_b}$$

Substituting for $i_b = i_e/(\beta + 1)$ where i_e is given by

$$i_e = \frac{v_i}{r_e + R_L}$$

we obtain

$$R_{\text{in}} = (\beta + 1)(r_e + R_L) \tag{7.129}$$

a result that we could have written directly using the resistance-reflection rule. Note that as expected the emitter follower takes the low load resistance and reflects it to the base side, where the signal source is, after increasing its value by a factor $(\beta + 1)$. It is this impedance transformation property of the emitter follower that makes it useful in connecting a low-resistance load to a high-resistance source, that is, to implement a buffer amplifier.

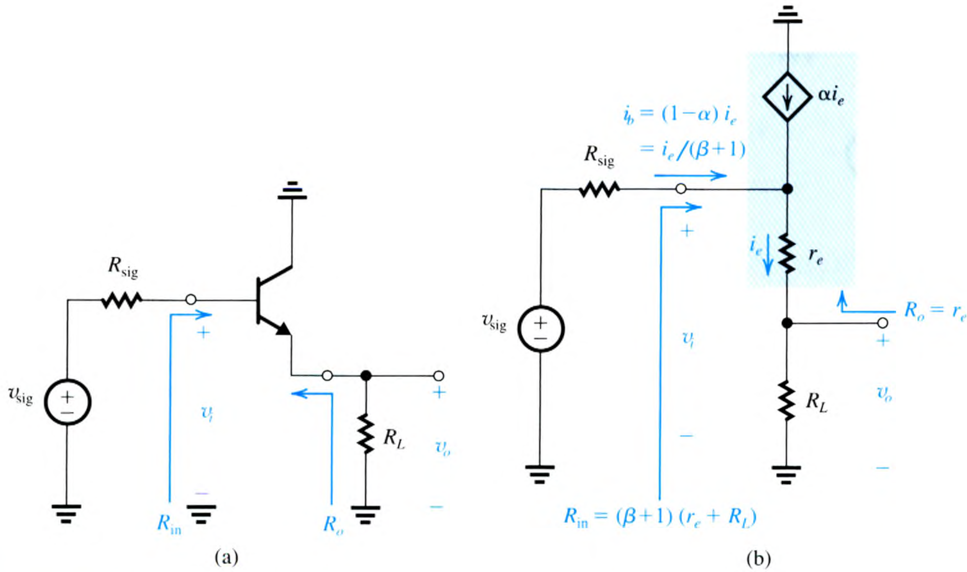


Figure 7.44 (**a**) Common-collector amplifier or emitter follower with the bias circuit omitted. (**b**) Equivalent circuit obtained by replacing the BJT with its T model.

The voltage gain A_v is given by

$$A_v \equiv \frac{v_o}{v_i} = \frac{R_L}{R_L + r_e} \tag{7.130}$$

Setting $R_L = \infty$ yields A_{vo},

$$A_{vo} = 1 \tag{7.131}$$

Thus, as expected, the open-circuit voltage gain of the emitter follower proper is one, which means that the signal voltage at the emitter *follows* that at the base, which is the origin of the name "emitter follower."

To determine R_o, refer to Fig. 7.44(b) and look back into the emitter (i.e., behind or excluding R_L) while setting $v_i = 0$ (i.e., grounding the base). You will see r_e of the BJT, thus

$$R_o = r_e \tag{7.132}$$

This result together with $A_{vo} = 1$ yields A_v in Eq. (7.130), thus confirming our earlier analysis.

We next determine the overall voltage gain G_v, as follows:

$$\frac{v_i}{v_{sig}} = \frac{R_{in}}{R_{in} + R_{sig}}$$
$$= \frac{(\beta+1)(r_e + R_L)}{(\beta+1)(r_e + R_L) + R_{sig}}$$
$$G_v \equiv \frac{v_o}{v_{sig}} = \frac{v_i}{v_{sig}} \times A_v$$

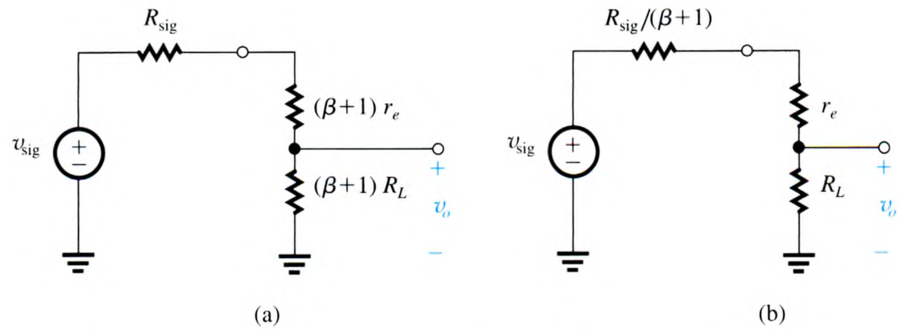


Figure 7.45 Simple equivalent circuits for the emitter follower obtained by (**a**) reflecting r_e and R_L to the base side, and (**b**) reflecting v_{sig} and R_{sig} to the emitter side. Note that the circuit in (**b**) can be obtained from that in (**a**) by simply dividing all resistances by $(\beta + 1)$.

Using the above expression for v_i/v_{sig} and substituting for A_v from Eq. (7.130) gives us

$$G_v = \frac{(\beta+1)R_L}{(\beta+1)R_L + (\beta+1)r_e + R_{sig}} \tag{7.133}$$

This equation indicates that the overall gain, though lower than one, can be close to one if $(\beta + 1)R_L$ is larger or comparable in value to R_{sig}. This again confirms the action of the emitter follower in delivering a large proportion of v_{sig} to a low-valued load resistance R_L even though R_{sig} can be much larger than R_L. The key point is that R_L is multiplied by $(\beta + 1)$ before it is "presented to the source." Figure 7.45(a) shows an equivalent circuit of the emitter follower obtained by simply reflecting r_e and R_L to the base side. The overall voltage gain $G_v \equiv v_o/v_{sig}$ can be determined directly and very simply from this circuit by using the voltage divider rule. The result is the expression for G_v already given in Eq. (7.133).

Dividing all resistances in the circuit of Fig. 7.45(a) by $\beta + 1$ does not change the voltage ratio v_o/v_{sig}. Thus we obtain another equivalent circuit, shown in Fig. 7.45(b), that we can use to determine $G_v \equiv v_o/v_{sig}$ of the emitter follower. A glance at this circuit reveals that it is simply the equivalent circuit obtained by reflecting v_{sig} and R_{sig} from the base side to the emitter side. In this reflection, v_{sig} does not change, but R_{sig} is divided by $\beta + 1$. Thus, we either reflect to the base side and obtain the circuit in Fig. 7.45(a) or we reflect to the emitter side and obtain the circuit in Fig. 7.45(b). From the latter, G_v can be found as

$$G_v \equiv \frac{v_o}{v_{sig}} = \frac{R_L}{R_L + r_e + R_{sig}/(\beta+1)} \tag{7.134}$$

Notice that this expression is the same as the one in Eq. (7.133) except for dividing both the numerator and denominator by $\beta + 1$.

The expression for G_v in Eq. (7.134) has an interesting interpretation: The emitter follower reduces R_{sig} by the factor $(\beta + 1)$ before "presenting it to the load resistance R_L": an impedance transformation that has the same buffering effect.

At this point it is important to note that although the emitter follower does not provide voltage gain it has a current gain of $\beta + 1$.

Thévenin Representation of the Emitter-Follower Output A more general representation of the emitter-follower output is shown in Fig. 7.46(a). Here G_{vo} is the overall open-circuit voltage gain that can be obtained by setting $R_L = \infty$ in the circuit of Fig. 7.45(b),

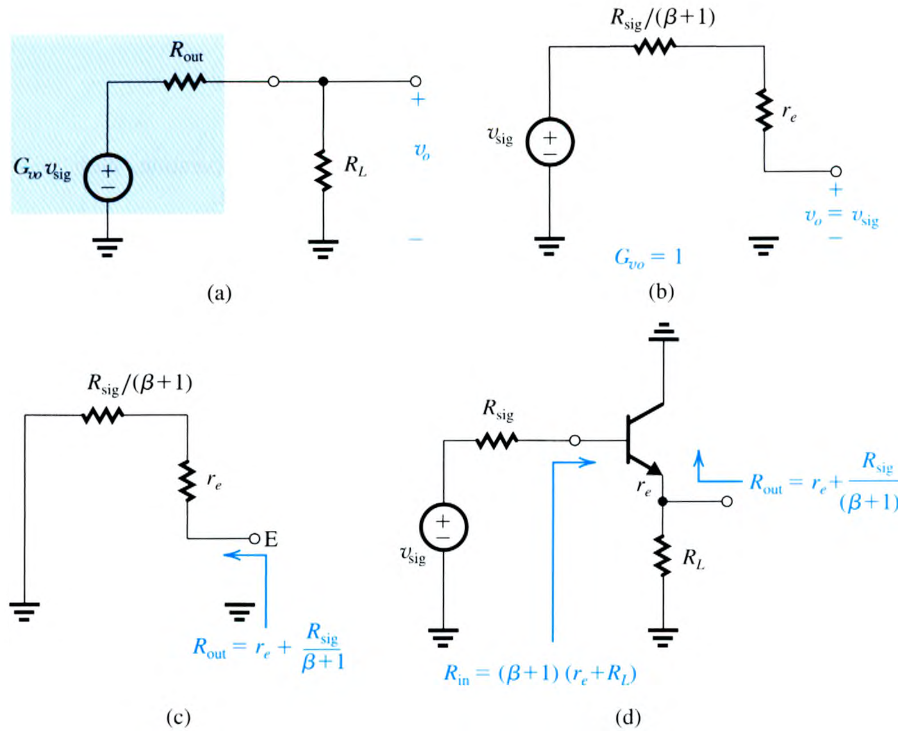


Figure 7.46 (**a**) Thévenin representation of the output of the emitter follower. (**b**) Obtaining G_{vo} from the equivalent circuit in Fig. 7.45(b). (**c**) Obtaining R_{out} from the equivalent circuit in Fig. 7.45(b) with v_{sig} set to zero. (**d**) The emitter follower with R_{in} and R_{out} determined simply by looking into the input and output terminals, respectively.

as illustrated in Fig. 7.46(b). The result is $G_{vo} = 1$. The output resistance R_{out} is *different* from R_o. To determine R_{out}, we set v_{sig} to zero (rather than setting v_i to zero). Again we can use the equivalent circuit in Fig. 7.45(b) to do this, as illustrated in Fig. 7.46(c). We see that

$$R_{out} = r_e + \frac{R_{sig}}{\beta+1} \tag{7.135}$$

Finally, in Fig. 7.46(d) we show the emitter-follower circuit together with its R_{in} and R_{out}. Observe that R_{in} is determined by reflecting r_e and R_L to the base side (by multiplying their values by $\beta+1$). To determine R_{out}, grab hold of the emitter and look backward while $v_{sig} = 0$. You will see r_e in series with R_{sig}, which, because it is in the base, must be divided by $(\beta+1)$.

We note that unlike the amplifier circuits we studied earlier, the emitter follower is *not* unilateral. This is manifested by the fact that R_{in} depends on R_L and R_{out} depends on R_{sig}.

Example 7.10

We need to design an emitter follower to implement the buffer amplifier of Fig. 7.47(a). Specify the required bias current I_E and the minimum value the transistor β must have. Determine the maximum

Example 7.10 *continued*

allowed value of v_{sig} if v_π is to be limited to 5 mV in order to obtain reasonably linear operation. With $v_{sig} = 200$ mV, find the signal voltage at the output if R_L is changed to 2 kΩ, and to 0.5 kΩ.

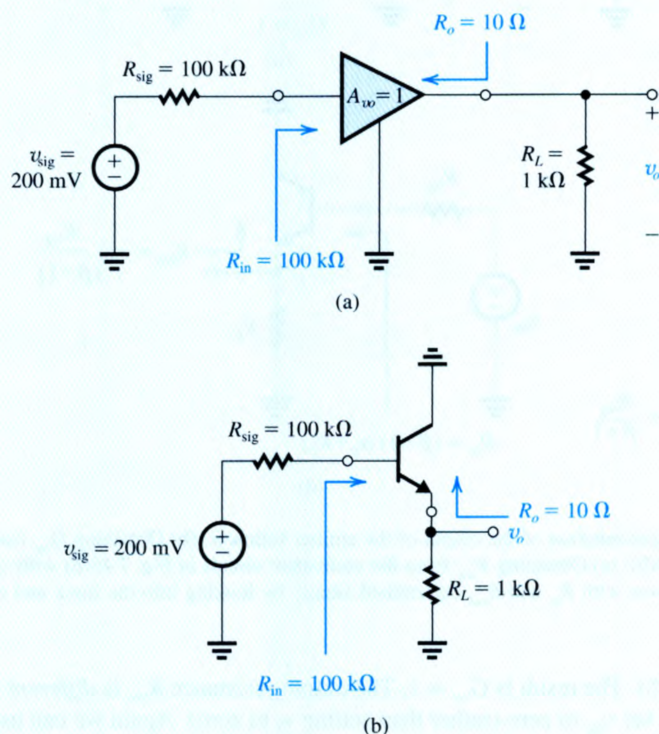


Figure 7.47 Circuit for Example 7.10.

Solution

The emitter-follower circuit is shown in Fig. 7.47(b). To obtain $R_o = 10\ \Omega$, we bias the transistor to obtain $r_e = 10\ \Omega$. Thus,

$$10\ \Omega = \frac{V_T}{I_E} = \frac{25\text{ mV}}{I_E}$$

$$\Rightarrow I_E = 2.5\text{ mA}$$

The input resistance R_{in} will be

$$R_{in} = (\beta + 1)(r_e + R_L)$$

$$100 = (\beta + 1)(0.01 + 1)$$

Thus, the BJT should have a β with a minimum value of 98. A higher β would obviously be beneficial.

We can find the overall voltage gain from

$$G_v \equiv \frac{v_o}{v_{sig}} = \frac{R_L}{R_L + r_e + \dfrac{R_{sig}}{(\beta+1)}}$$

Assuming $\beta = 100$, the value of G_v is

$$G_v = 0.5$$

Thus when $v_{sig} = 200$ mV, the signal at the output will be 100 mV. Since the 100 mV appears across the 1-kΩ load, we can find the signal across the base–emitter junction from

$$v_\pi = \frac{v_o}{R_L} \times r_e$$
$$= \frac{100}{1000} \times 10 = 1 \text{ mV}$$

If $\hat{v}_\pi = 5$ mV then v_{sig} can be increased by a factor of 5, resulting in $\hat{v}_{sig} = 1$ V.

To obtain v_o as the load is varied, we use the Thévenin equivalent of the emitter follower, shown in Fig. 7.46(a), with $G_{vo} = 1$ and

$$R_{out} = \frac{R_{sig}}{\beta+1} + r_e = \frac{100}{101} + 0.01 = 1 \text{ k}\Omega$$

to obtain

$$v_o = v_{sig} \frac{R_L}{R_L + R_{out}}$$

For $R_L = 2$ kΩ,

$$v_o = 200 \text{ mV} \times \frac{2}{2+1} = 133.3 \text{ mV}$$

and for $R_L = 0.5$ kΩ,

$$v_o = 200 \text{ mV} \times \frac{0.5}{0.5+1} = 66.7 \text{ mV}$$

EXERCISE

7.30 An emitter follower uses a transistor with $\beta = 100$ and is biased at $I_C = 5$ mA. It operates between a source with a resistance of 10 kΩ and a load of 1 kΩ. Find R_{in}, G_{vo}, R_{out}, and G_v. What is the peak amplitude of v_{sig} that results in v_π having a peak amplitude of 5 mV? Find the resulting peak amplitude at the output.

Ans. 101.5 kΩ; 1 V/V; 104 Ω; 0.91 V/V; 1.1 V; 1 V

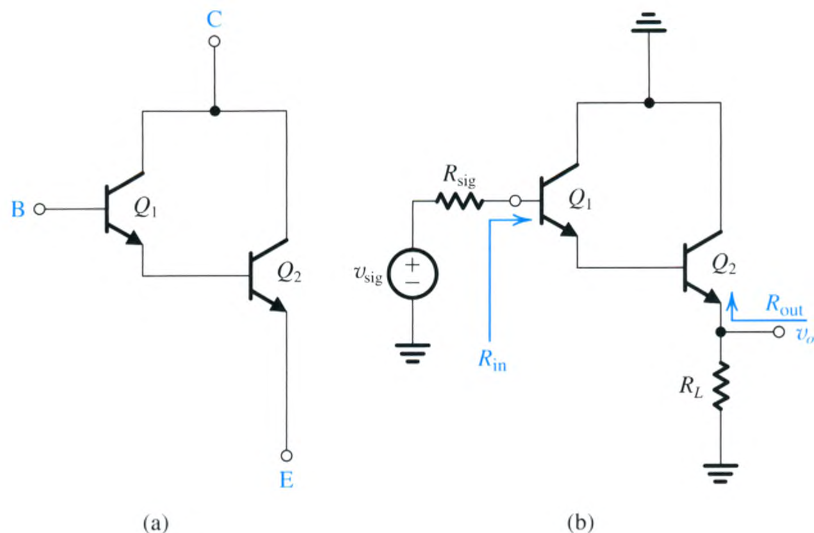


Figure 7.48 (**a**) The Darlington configuration. (**b**) Voltage follower using the Darlington configuration.

Enhancing the Emitter-Follower: The Darlington Configuration From the above we see that the performance of the emitter follower can be improved by increasing the value of β. It follows that we can obtain an enhanced emitter follower by placing two BJTs in cascade, as shown in Fig. 7.48(a). This is known as the **Darlington configuration**, and the pair of cascaded transistors as the **Darlington pair**. The Darlington pair can be thought of as a composite transistor with $\beta = \beta_1\beta_2$. It can be used to implement a high-performance voltage follower, as shown in Fig. 7.48(b).

EXERCISE

7.31 For the Darlington voltage follower in Fig. 7.48(b), show that:

$$R_{in} = (\beta_1 + 1)[r_{e1} + (\beta_2 + 1)(r_{e2} + R_L)]$$

$$R_{out} = R_L \| r_{e2} + \frac{r_{e1} + [R_{sig}/(\beta_1 + 1)]}{\beta_2 + 1}$$

$$\frac{v_o}{v_{sig}} = \frac{R_L}{R_L + r_{e2} + [r_{e1} + R_{sig}/(\beta_1 + 1)]/(\beta_2 + 1)}$$

Evaluate R_{in}, R_{out}, and v_o/v_{sig} for the case $I_{E2} = 5$ mA, $\beta_1 = \beta_2 = 100$, $R_E = 1$ kΩ, and $R_{sig} = 100$ kΩ.

Ans. 10.3 MΩ; 20 Ω; 0.98 V/V

7.3.7 Summary Tables and Comparisons

For easy reference and to enable comparisons, we present the formulas for determining the characteristic parameters for the various configurations of MOSFET and BJT amplifiers in Tables 7.4 and 7.5, respectively; you will find these in the Summary Tables supplement found online at www.oup.com/he/sedra-smith8e. In addition to the remarks made throughout this section about the characteristics and applications of the various configurations, we make the following concluding points:

1. MOS amplifiers provide much higher, ideally infinite input resistances (except, of course, for the CG configuration). This is a definite advantage over BJT amplifiers.
2. BJTs exhibit higher g_m values than MOSFETs, resulting in higher gains.
3. For discrete-circuit amplifiers—that is, those assembled from discrete components on a printed-circuit board (PCB)—the BJT remains the device of choice. This is because discrete BJTs are much easier to handle physically than discrete MOSFETs and, more important, a wide variety of discrete BJTs are available commercially. The rest of this chapter is concerned with discrete-circuit amplifiers.
4. Integrated-circuit (IC) amplifiers mostly use MOSFETs, although BJTs are used in certain niche areas. Chapters 8 to 13 are mainly concerned with IC amplifiers.
5. The CS and CE configurations are best suited for realizing the bulk of the gain required in an amplifier. Depending on the magnitude of the gain required, either a single stage or a cascade of two or three stages can be used.
6. Including a resistance R_s in the source of the CS amplifier (a resistance R_e in the emitter of the CE amplifier) provides a number of performance improvements at the expense of gain reduction.
7. The low input resistance of the CG and CB amplifiers makes them useful only in specific applications. As we will see in Chapter 10, these two configurations exhibit a much better high-frequency response than that available from the CS and CE amplifiers. This makes them useful as high-frequency amplifiers, especially when combined with the CS or CE circuit. We shall study one such combination in Chapter 8.
8. The source follower (emitter follower) can be used as a voltage buffer for connecting a high-resistance source to a low-resistance load, and as the output stage in a multistage amplifier, where its purpose is to equip the amplifier with a low output resistance.

7.3.8 When and How to Include the Output Resistance r_o

So far we have been neglecting the output resistance r_o of the MOSFET and the BJT. We have done this for two reasons:

1. To keep things simple and focus attention on the significant features of each of the basic configurations, and
2. Because our main interest in this chapter is discrete-circuit design, where the circuit resistances (e.g., R_C, R_D, and R_L) are usually much smaller than r_o.

Nevertheless, in some instances it is relatively easy to include r_o in the analysis. Specifically:

1. In the CS and CE amplifiers, we can see that r_o of the transistor appears in parallel with R_D and R_C, respectively, and can be simply included in the corresponding formulas by replacing R_D with $(R_D \| r_o)$ and R_C with $(R_C \| r_o)$. The effect will be a reduction in the magnitude of gain, of perhaps 5% to 10%.
2. In the source and emitter followers, we can see that the transistor r_o appears in parallel with R_L and can be taken into account by replacing R_L in the corresponding formulas with $(R_L \| r_o)$. Here too, the effect of taking r_o into account is a small reduction in gain. More significant, however, taking r_o into account reduces the open-circuit voltage gain A_{vo} from unity to

$$A_{vo} = \frac{r_o}{r_o + (1/g_m)} \tag{7.136}$$

There are configurations in which taking r_o into account complicates the analysis considerably. These are the CS (CE) amplifier with a source (emitter) resistance, and the CG (CB) amplifier. Fortunately, for discrete implementation of these configurations, the effect of neglecting r_o is usually small (which can be verified by computer simulation).

Finally, a very important point: *In the analysis and design of IC amplifiers, r_o must always be taken into account.* This is because, as we will see in the next chapter, all the circuit resistances are of the same order of magnitude as r_o; neglecting r_o can result in completely erroneous results.

LEE DE FOREST—A FATHER OF THE ELECTRONICS AGE

In 1906 self-employed inventor Lee de Forest (1873–1961) created a three-terminal vacuum tube; it was the first electronic amplifier of weak signals. The device was known initially as the de Forest valve. The patent filed in 1907, however, used the name Audion, with the "-ion" indicating that the device was not completely evacuated. By 1919, engineers had realized that complete evacuation of internal gases produced a more reliable device.

De Forest's first amplifier became known as the vacuum tube triode. Through its impact on radio, telephony, motion picture sound, and television, this invention, one of de Forest's 180 patents, is credited with introducing the electronics age. The vacuum tube, in a variety of types, remained the device for implementing amplifiers until the appearance of transistors in the early 1950s.

7.4 Biasing

As we discussed in Section 7.1, an essential step in the design of a transistor amplifier is establishing an appropriate dc operating point for the transistor. This is known as biasing or bias design. In this section, we study the biasing methods commonly used in discrete-circuit amplifiers. Biasing of integrated-circuit amplifiers will be studied in Chapter 8.

Bias design aims to establish a dc current in the drain (collector) that is predictable and insensitive to variations in temperature and to the large variations in parameter values between devices of the same type. For instance, discrete BJTs belonging to the same manufacturer's part number can have β values that range, say, from 50 to 150. Nevertheless, the bias design for an amplifier using this particular transistor type may specify that the dc

collector current shall always be within, say, ±10% of the nominal value of, say, 1 mA. We can make a similar statement about the desired insensitivity of the dc drain current to the wide variations encountered in V_t of discrete MOSFETs.

A second consideration in bias design is locating the dc operating point in the active region of operation of the transistor to obtain high voltage gain while allowing for the required output signal swing without letting the transistor leave the active region at any time (in order to avoid nonlinear distortion). We discussed this point in Section 7.1.7.

Although we will consider the biasing of MOSFET and BJT amplifiers separately, the resulting circuits are very similar. Also, we will see that good bias designs incorporate a feedback mechanism that works to keep the dc bias point as constant as possible.

In order to keep matters simple and focus our attention on significant issues, we will neglect the Early effect; that is, assume $\lambda = 0$ or $V_A = \infty$. This is certainly allowed in initial designs of discrete circuits. Of course, the design can be fine-tuned at a later point with the assistance of a circuit-simulation program such as SPICE.

7.4.1 The MOSFET Case

Biasing by Fixing V_{GS} The most straightforward approach to biasing a MOSFET is to fix its gate-to-source voltage V_{GS} to the value required[5] to provide the desired I_D. We can derive the voltage value from the power-supply voltage V_{DD} by using an appropriate voltage divider, as shown in Fig. 7.49(a). However, no matter how the voltage V_{GS} is generated, this is *not* a good approach to biasing a MOSFET. To understand why, recall that

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_t)^2$$

and note that the values of the threshold voltage V_t, the oxide-capacitance C_{ox}, and (to a lesser extent) the transistor aspect ratio W/L vary widely among devices of supposedly the same size and type. This is certainly the case for discrete devices, in which large spreads in the values of these parameters occur among devices of the same manufacturer's part number. The spread can also be large in integrated circuits, especially among devices fabricated on different wafers and certainly between different batches of wafers. Furthermore, both V_t and μ_n depend on temperature, with the result that if we fix the value of V_{GS}, the drain current I_D becomes very much temperature dependent.

To emphasize the point that biasing by fixing V_{GS} is not a good technique, we show in Fig. 7.49(b) two i_D–v_{GS} characteristic curves representing extreme values in a batch of MOSFETs of the same type. Observe that for the fixed value of V_{GS}, the resultant spread in the values of the drain current can be substantial.

Biasing by Fixing V_G and Connecting a Resistance in the Source An excellent biasing technique for discrete MOSFET circuits consists of fixing the dc voltage at the gate, V_G, and connecting a resistance in the source lead, as shown in Fig. 7.50(a). For this circuit,

$$V_G = V_{GS} + R_S I_D \tag{7.137}$$

[5]That is indeed what we were doing in Section 7.1. However, the amplifier circuits studied there were conceptual ones, not actual practical circuits. Our purpose in this section is to study the latter.

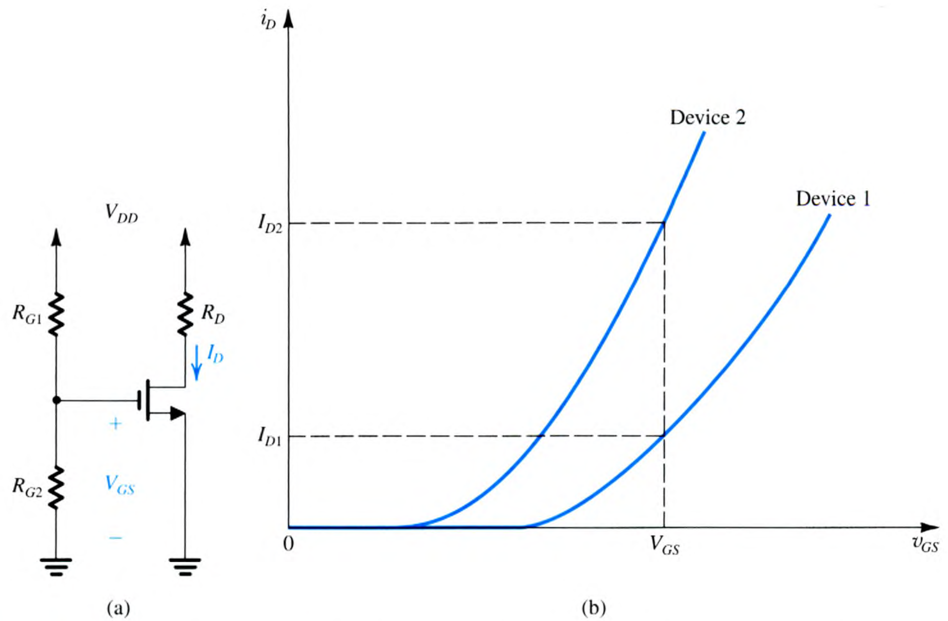


Figure 7.49 (**a**) Biasing the MOSFET with a constant V_{GS} generated from V_{DD} using a voltage divider (R_{G1},R_{G2}). (**b**) The use of fixed bias (constant V_{GS}) can result in a large variability in the value of I_D. Devices 1 and 2 represent extremes among units of the same type.

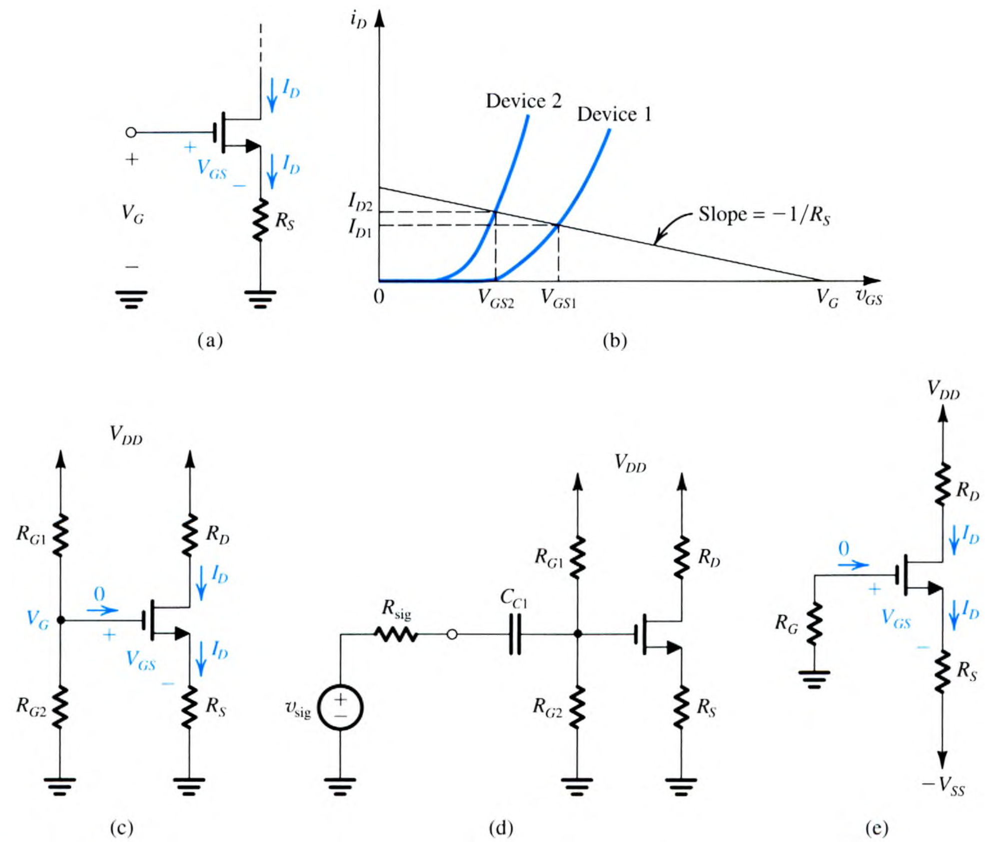


Figure 7.50 Biasing using a fixed voltage at the gate, V_G, and a resistance in the source lead, R_S: **(a)** basic arrangement; **(b)** reduced variability in I_D; **(c)** practical implementation using a single supply; **(d)** coupling of a signal source to the gate using a capacitor C_{C1}; **(e)** practical implementation using two supplies.

Now, if V_G is much greater than V_{GS}, I_D will be mostly determined by the values of V_G and R_S. However, even if V_G is not much larger than V_{GS}, resistor R_S provides *negative feedback*, which acts to stabilize the value of the bias current I_D. To see how this comes about, consider what happens when I_D increases for whatever reason. Equation (7.137) indicates that since V_G is constant, V_{GS} will have to decrease. This in turn results in a decrease in I_D, a change that is opposite to what we initially assumed. Thus the action of R_S works to keep I_D as constant as possible.[6]

Figure 7.50(b) illustrates the effectiveness of this biasing scheme. Here too, we show the i_D–v_{GS} characteristics for two devices that represent the extremes of a batch of MOSFETs. Superimposed on the device characteristics is a straight line that represents the constraint imposed by the bias circuit—namely, Eq. (7.137). The intersection of this straight line with the i_D–v_{GS} characteristic curve provides the coordinates (I_D and V_{GS}) of the bias point. Notice that compared to the case of fixed V_{GS}, here the variability obtained in I_D is much smaller. Also, note that the variability decreases as V_G and R_S are made larger (thus providing a bias line that is less steep).

Two possible practical discrete implementations of this bias scheme are shown in Figs. 7.50(c) and (e). The circuit in Fig. 7.50(c) uses one power-supply V_{DD} and derives V_G through a voltage divider (R_{G1}, R_{G2}). Since $I_G = 0$, R_{G1} and R_{G2} can be selected to be very large (in the megohm range), allowing the MOSFET to present a large input resistance to a signal source that may be connected to the gate through a coupling capacitor, as shown in Fig. 7.50(d). Here capacitor C_{C1} blocks dc and thus allows us to couple the signal v_{sig} to the amplifier input without disturbing the MOSFET dc bias point. The selected value of C_{C1} should be large enough to approximate a short circuit at all signal frequencies of interest. We will study capacitively coupled MOSFET amplifiers, which are suitable only in discrete-circuit design, in Section 7.5. Finally, note that in the circuit of Fig. 7.50(c), resistor R_D is selected to be as large as possible to obtain high gain but small enough to allow for the desired signal swing at the drain while keeping the MOSFET in saturation at all times.

When two power supplies are available, as is often the case, we can use the somewhat simpler bias arrangement of Fig. 7.50(e). This circuit is an implementation of Eq. (7.137), with V_G replaced by V_{SS}. Resistor R_G establishes a dc ground at the gate and presents a high input resistance to a signal source that may be connected to the gate through a coupling capacitor.

Example 7.11

We need to design the circuit of Fig. 7.50(c) to establish a dc drain current $I_D = 0.5$ mA. The MOSFET is specified to have $V_t = 1$ V and $k_n' W/L = 1$ mA/V^2. For simplicity, neglect the Early effect (i.e., assume $\lambda = 0$). Use a power-supply $V_{DD} = 15$ V. Calculate the percentage change in the value of I_D when the MOSFET is replaced with another unit having the same $k_n' W/L$ but $V_t = 1.5$ V.

[6]The action of R_S in stabilizing the value of the bias current I_D is not unlike that of the resistance R_s, which we included in the source lead of a CS amplifier in Section 7.3.4. In the latter case also, R_s works to reduce the change in i_D with the result that the amplifier gain is reduced.

Example 7.11 *continued*

Solution

As a rule of thumb for designing this classical biasing circuit, we choose R_D and R_S to provide one-third of the power-supply voltage V_{DD} as a drop across each of R_D, the transistor (i.e., V_{DS}), and R_S. For $V_{DD} = 15$ V, this choice makes $V_D = +10$ V and $V_S = +5$ V. Now, since I_D is required to be 0.5 mA, we can find the values of R_D and R_S as follows:

$$R_D = \frac{V_{DD} - V_D}{I_D} = \frac{15 - 10}{0.5} = 10 \text{ k}\Omega$$

$$R_S = \frac{V_S}{R_S} = \frac{5}{0.5} = 10 \text{ k}\Omega$$

The required value of V_{GS} can be determined by first calculating the overdrive voltage V_{OV} from

$$I_D = \frac{1}{2} k_n'(W/L) V_{OV}^2$$

$$0.5 = \frac{1}{2} \times 1 \times V_{OV}^2$$

which gives $V_{OV} = 1$ V, and thus,

$$V_{GS} = V_t + V_{OV} = 1 + 1 = 2 \text{ V}$$

Now, since $V_S = +5$ V, V_G must be

$$V_G = V_S + V_{GS} = 5 + 2 = 7 \text{ V}$$

To establish this voltage at the gate, we can select $R_{G1} = 8$ MΩ and $R_{G2} = 7$ MΩ. The final circuit is shown in Fig. 7.51. Observe that the dc voltage at the drain (+10 V) allows for a positive signal swing of +5 V (i.e., up to V_{DD}) and a negative signal swing of 4 V [i.e., down to $(V_G - V_t)$].

If the NMOS transistor is replaced with another having $V_t = 1.5$ V, we can find the new value of I_D as follows:

$$I_D = \frac{1}{2} \times 1 \times (V_{GS} - 1.5)^2 \qquad (7.138)$$

$$V_G = V_{GS} + I_D R_S$$

$$7 = V_{GS} + 10 I_D \qquad (7.139)$$

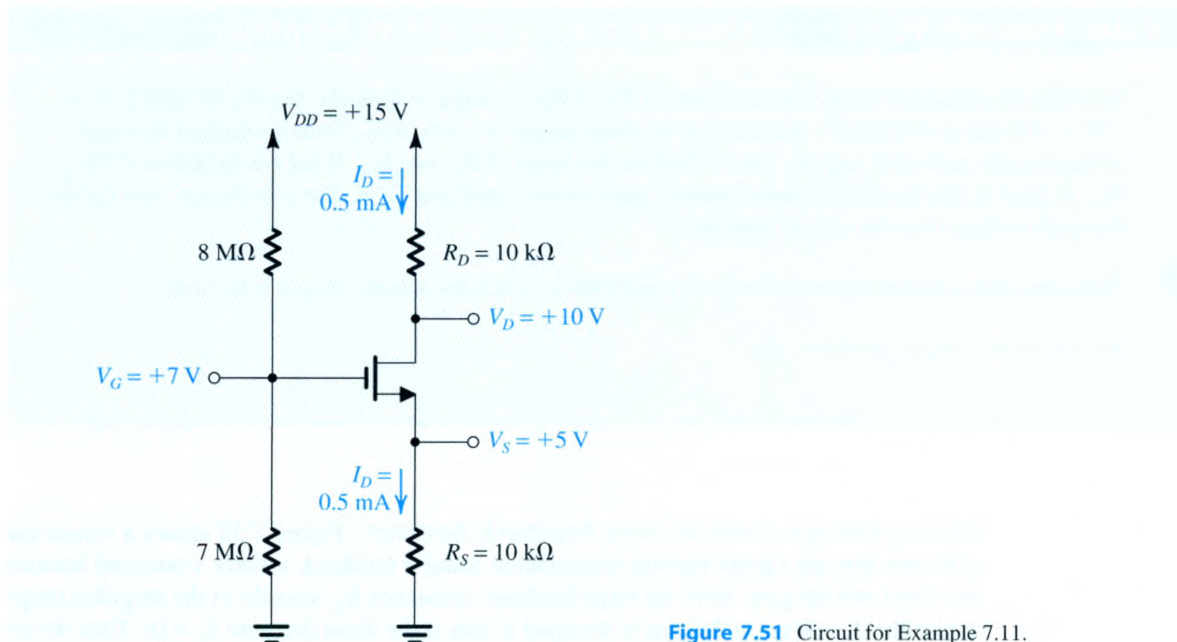


Figure 7.51 Circuit for Example 7.11.

Solving Eqs. (7.138) and (7.139) together yields

$$I_D = 0.455 \text{ mA}$$

Thus the change in I_D is

$$\Delta I_D = 0.455 - 0.5 = -0.045 \text{ mA}$$

which is $\dfrac{-0.045}{0.5} \times 100 = -9\%$ change.

EXERCISES

7.32 Consider the MOSFET in Example 7.11 when fixed-V_{GS} bias is used. Find the required value of V_{GS} to establish a dc bias current $I_D = 0.5$ mA. Recall that the device parameters are $V_t = 1$ V, $k_n' W/L =$ 1 mA/V^2, and $\lambda = 0$. What is the percentage change in I_D when the transistor is replaced with another having $V_t = 1.5$ V?

Ans. $V_{GS} = 2$ V; –75%

D7.33 Design the circuit of Fig. 7.50(e) to operate at a dc drain current of 0.5 mA and $V_D = +2$ V. Let $V_t =$ 1 V, $k_n' W/L = 1$ mA/V^2, $\lambda = 0$, $V_{DD} = V_{SS} = 5$ V. Use standard 5% resistor values (see Appendix J), and give the resulting values of I_D, V_D, and V_S.

Ans. $R_D = R_S = 6.2$ kΩ; $I_D = 0.49$ mA, $V_S = -1.96$ V, and $V_D = +1.96$ V. R_G can be selected in the range of 1 MΩ to 10 MΩ.

Video Example VE 7.2 Classical Biasing of a MOS Amplifier

Consider the classical biasing scheme shown in Fig. 7.50(c), using a 9-V supply. For the MOSFET, $V_t =$ 1 V, $\lambda = 0$, and $k_n = 2$ mA/V^2. Arrange that the drain current is 1 mA, with about one-third of the supply voltage across each of R_S and R_D. Use 22 MΩ for the larger of R_{G1} and R_{G2}. What are the values of R_{G1}, R_{G2}, R_S, and R_D that you have chosen? Specify them to two significant digits. For your design, how far is the drain voltage from the edge of saturation?

VE 7.2

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 7.92

Biasing Using a Drain-to-Gate Feedback Resistor Figure 7.52 shows a simple and effective discrete-circuit biasing arrangement using a feedback resistor connected between the drain and the gate. Here the large feedback resistance R_G (usually in the megohm range) forces the dc voltage at the gate to be equal to that at the drain (because $I_G = 0$). Thus we can write

$$V_{GS} = V_{DS} = V_{DD} - R_D I_D$$

which can be rewritten

$$V_{DD} = V_{GS} + R_D I_D \tag{7.140}$$

This is identical in form to Eq. (7.137), which describes the operation of the bias scheme discussed above [that in Fig. 7.50(a)]. Thus, here too, if I_D for some reason changes, say increases, then Eq. (7.140) indicates that V_{GS} must decrease. The decrease in V_{GS} in turn causes a decrease in I_D, a change that is opposite in direction to the one we originally assumed. Thus the negative feedback or degeneration provided by R_G works to keep the value of I_D as constant as possible.

The circuit of Fig. 7.52 can be used as an amplifier by applying the input voltage signal to the gate via a coupling capacitor so as not to disturb the dc bias conditions already established. The amplified output signal at the drain can be coupled to another part of the circuit, again via a capacitor. We considered such an amplifier circuit in Section 7.2 (Example 7.3).

EXERCISE

D7.34 Design the circuit in Fig. 7.52 to operate at a dc drain current of 0.5 mA. Assume $V_{DD} = +5$ V, $k_n' W/L = 1$ mA/V^2, $V_t = 1$ V, and $\lambda = 0$. Use a standard 5% resistance value for R_D (see Appendix J), and give the actual values obtained for I_D and V_D.

Ans. $R_D = 6.2$ kΩ; $I_D \simeq 0.49$ mA; $V_D \simeq 1.96$ V

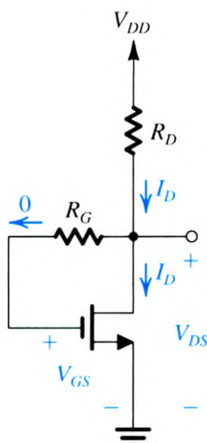


Figure 7.52 Biasing the MOSFET using a large drain-to-gate feedback resistance, R_G.

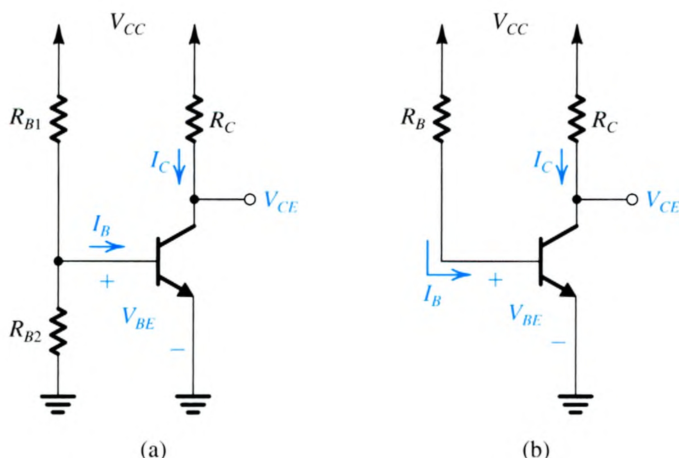


Figure 7.53 Two obvious schemes for biasing the BJT: **(a)** by fixing V_{BE}; **(b)** by fixing I_B. Both result in wide variations in I_C and hence in V_{CE} and therefore are considered to be "bad." Neither scheme is recommended.

7.4.2 The BJT Case

Before presenting the "good" biasing schemes, we should point out why two obvious arrangements are *not* good. First, attempting to bias the BJT by fixing the voltage V_{BE} by, for instance, using a voltage divider across the power supply V_{CC}, as shown in Fig. 7.53(a), is not a viable approach: The very sharp exponential relationship $i_C - v_{BE}$ means that any small and inevitable differences in V_{BE} from the desired value will result in large differences in I_C and in V_{CE}. Second, biasing the BJT by establishing a constant current in the base, as shown in Fig. 7.53(b), where $I_B \simeq (V_{CC} - 0.7)/R_B$, is also not recommended. Here the typically large variations in the value of β among units of the same device type will result in correspondingly large variations in I_C and hence in V_{CE}.

The Classical Discrete-Circuit Bias Arrangement Figure 7.54(a) shows the arrangement most commonly used for biasing a discrete-circuit transistor amplifier if only a single power supply is available. The technique consists of supplying the base of the transistor with

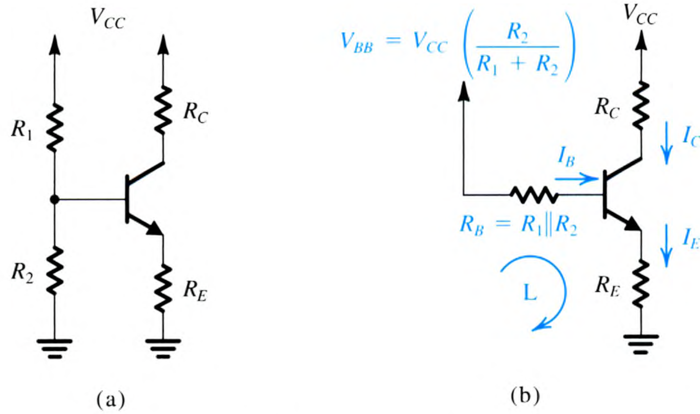


Figure 7.54 Classical biasing for BJTs using a single power supply: (**a**) circuit; (**b**) circuit with the voltage divider supplying the base replaced with its Thévenin equivalent.

a fraction of the supply voltage V_{CC} through the voltage divider R_1, R_2. In addition, a resistor R_E is connected to the emitter. This circuit is very similar to one we used for the MOSFET [Fig. 7.50(c)]. Here, however, the design must take into account the finite base current.

Figure 7.54(b) shows the same circuit with the voltage-divider network replaced by its Thévenin equivalent,

$$V_{BB} = \frac{R_2}{R_1 + R_2} V_{CC} \tag{7.141}$$

$$R_B = \frac{R_1 R_2}{R_1 + R_2} \tag{7.142}$$

We can find the current I_E by writing a Kirchhoff loop equation for the base–emitter–ground loop, labeled L, and substituting $I_B = I_E/(\beta + 1)$:

$$I_E = \frac{V_{BB} - V_{BE}}{R_E + R_B/(\beta + 1)} \tag{7.143}$$

To make I_E insensitive to temperature and β variation,[7] we design the circuit to satisfy the following two constraints:

$$V_{BB} \gg V_{BE} \tag{7.144}$$

$$R_E \gg \frac{R_B}{\beta + 1} \tag{7.145}$$

Condition (7.144) ensures that small variations in V_{BE} ($\simeq 0.7$ V) will be swamped by the much larger V_{BB}. There is a limit, however, on how large V_{BB} can be: For a given value of the supply voltage V_{CC}, the higher the value we use for V_{BB}, the lower will be the sum of voltages across R_C and the collector–base junction (V_{CB}). On the other hand, we want the voltage across R_C to be large in order to obtain high voltage gain and large signal swing (before transistor cutoff). We also want V_{CB} (or V_{CE}) to be large, to provide a large signal swing (before transistor

[7] Bias design seeks to stabilize either I_E or I_C since $I_C = \alpha I_E$ and α varies very little. That is, a stable I_E will result in an equally stable I_C, and vice versa.

saturation). Thus, as is the case in any design, we have a set of conflicting requirements, and the solution must be a trade-off. As a rule of thumb, we design for V_{BB} about $\frac{1}{3}V_{CC}$, V_{CB} (or V_{CE}) about $\frac{1}{3}V_{CC}$, and $I_C R_C$ about $\frac{1}{3}V_{CC}$.

Condition (7.145) makes I_E insensitive to variations in β and could be satisfied by selecting small R_B. This in turn is achieved by using low values for R_1 and R_2. Lower values for R_1 and R_2, however, will mean a higher current drain from the power supply, and will result in a lowering of the input resistance of the amplifier (if the input signal is coupled to the base),[8] which is the trade-off involved in this part of the design. We should note that condition (7.145) means that we want to make the base voltage independent of the value of β and determined solely by the voltage divider. This will obviously be satisfied if the current in the divider is made much larger than the base current. Typically we select R_1 and R_2 such that their current is in the range of I_E to $0.1I_E$.

We can gain further insight into the mechanism by which the bias arrangement of Fig. 7.54(a) stabilizes the dc emitter (and hence collector) current by considering the feedback action provided by R_E. Let us say that for some reason the emitter current increases. The voltage drop across R_E, and hence V_E, will increase correspondingly. Now, if the base voltage is determined primarily by the voltage divider R_1, R_2, which is the case if R_B is small, it will remain constant, and the increase in V_E will result in a corresponding decrease in V_{BE}. This in turn reduces the collector (and emitter) current, a change opposite to what we had originally assumed. Thus R_E provides a *negative feedback* action that stabilizes the bias current. This should remind you of the resistance R_e that we included in the emitter lead of the CE amplifier in Section 7.3.4. Also, the feedback action of R_E in the circuit of Fig. 7.54(a) is similar to the feedback action of R_S in the circuit of Fig. 7.50(c). We will study negative feedback formally in Chapter 11.

Example 7.12

We want to design the bias circuit of the amplifier in Fig. 7.54(a) to establish a current $I_E = 1$ mA using a power supply $V_{CC} = +12$ V. The transistor is specified to have a nominal β value of 100.

Solution

We will follow the rule of thumb mentioned above and allocate one-third of the supply voltage to the voltage drop across R_2 and another one-third to the voltage drop across R_C, leaving one-third for possible negative signal swing at the collector. Thus,

$$V_B = +4 \text{ V} \qquad V_E = 4 - V_{BE} \simeq 3.3 \text{ V}$$

and R_E is found from

$$R_E = \frac{V_E}{I_E} = \frac{3.3 \text{ V}}{1 \text{ mA}} = 3.3 \text{ k}\Omega$$

[8] If the input signal is coupled to the transistor base, the two bias resistances R_1 and R_2 effectively appear in parallel between the base and ground. Thus, low values for R_1 and R_2 will result in lowering R_{in}.

Example 7.12 *continued*

From the discussion above, we select a voltage-divider current of $0.1I_E = 0.1 \times 1 = 0.1$ mA. Neglecting the base current, we find

$$R_1 + R_2 = \frac{12}{0.1} = 120\text{ k}\Omega$$

and

$$\frac{R_2}{R_1 + R_2}V_{CC} = 4\text{ V}$$

Thus $R_2 = 40\text{ k}\Omega$ and $R_1 = 80\text{ k}\Omega$.

At this point, we want to find a more accurate estimate for I_E, taking into account the nonzero base current. Using Eq. (7.143),

$$I_E = \frac{4 - 0.7}{3.3(\text{k}\Omega) + \frac{(80\|40)(\text{k}\Omega)}{101}} = 0.93\text{ mA}$$

This is quite a bit lower than 1 mA, the value we are aiming for. It is easy to see from the above equation that a simple way to restore I_E to its nominal value would be to reduce R_E from 3.3 kΩ by the magnitude of the second term in the denominator (0.267 kΩ). Thus a more suitable value for R_E in this case would be $R_E = 3\text{ k}\Omega$, which results in $I_E = 1.01\text{ mA} \simeq 1\text{ mA}$.[9]

We should note that if we are willing to draw a higher current from the power supply and to accept a lower input resistance for the amplifier, then we may use a voltage-divider current equal, say, to I_E (i.e., 1 mA), resulting in $R_1 = 8\text{ k}\Omega$ and $R_2 = 4\text{ k}\Omega$. We will refer to the circuit using these latter values as design 2, for which the actual value of I_E using the initial value of R_E of 3.3 kΩ will be

$$I_E = \frac{4 - 0.7}{3.3 + 0.027} = 0.99 \simeq 1\text{ mA}$$

In this case, design 2, we need not change the value of R_E.

Finally, we can find the value of R_C from

$$R_C = \frac{12 - V_C}{I_C}$$

Substituting $I_C = \alpha I_E = 0.99 \times 1 = 0.99\text{ mA} \simeq 1\text{ mA}$ results, for both designs, in

$$R_C = \frac{12 - 8}{1} = 4\text{ k}\Omega$$

[9]Although reducing R_E restores I_E to the design value of 1 mA, it does not solve the problem of the dependence of the value of I_E on β. See Exercise 7.35.

EXERCISE

7.35 For design 1 in Example 7.12, calculate the expected range of I_E if the transistor used has β in the range of 50 to 150. Express the range of I_E as a percentage of the nominal value ($I_E \simeq 1$ mA) obtained for $\beta = 100$. Repeat for design 2.

Ans. For design 1: 0.94 mA to 1.04 mA, a 10% range; for design 2: 0.984 mA to 0.995 mA, a 1.1% range.

A Two-Power-Supply Version of the Classical Bias Arrangement A somewhat simpler bias arrangement is possible if two power supplies are available, as shown in Fig. 7.55. Writing a loop equation for the loop labeled L gives

$$I_E = \frac{V_{EE} - V_{BE}}{R_E + R_B/(\beta + 1)} \tag{7.146}$$

This equation is identical to Eq. (7.143) except for V_{EE} replacing V_{BB}. Thus the two constraints of Eqs. (7.144) and (7.145) apply here as well. Note that if the transistor is to be used with the base grounded (i.e., in the common-base configuration), then R_B can be eliminated altogether. On the other hand, if the input signal is to be coupled to the base, then R_B is needed. We will study complete circuits of the various BJT amplifier configurations in Section 7.5. Finally, notice that the circuit in Fig. 7.55 is the counterpart of the MOS circuit in Fig. 7.50(e).

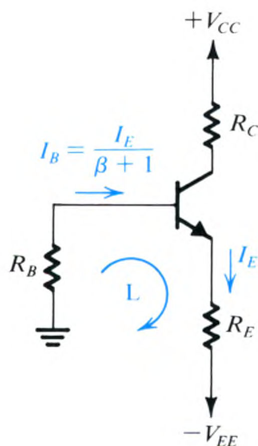


Figure 7.55 Biasing the BJT using two power supplies. Resistor R_B is needed only if the signal is to be capacitively coupled to the base. Otherwise, the base can be connected directly to ground, or to a grounded signal source, resulting in almost total β-independence of the bias current.

EXERCISE

D7.36 The bias arrangement of Fig. 7.55 is to be used for a common-base amplifier. Design the circuit to establish a dc emitter current of 1 mA and provide the highest possible voltage gain while allowing for a signal swing at the collector of ±2 V. Use +10-V and –5-V power supplies.
Ans. $R_B = 0$; $R_E = 4.3\ \text{k}\Omega$; $R_C = 8.4\ \text{k}\Omega$

Biasing Using a Collector-to-Base Feedback Resistor In the BJT case, there is a counterpart to the MOSFET circuit of Fig. 7.52. Figure 7.56(a) shows such a simple but effective biasing arrangement that is suitable for common-emitter amplifiers. The circuit uses a resistor R_B connected between the collector and the base. Resistor R_B provides negative feedback, which helps to stabilize the bias point of the BJT.

Analysis of the circuit is shown in Fig. 7.56(b), from which we can write

$$V_{CC} = I_E R_C + I_B R_B + V_{BE}$$
$$= I_E R_C + \frac{I_E}{\beta + 1} R_B + V_{BE}$$

Thus the emitter bias current is given by

$$I_E = \frac{V_{CC} - V_{BE}}{R_C + R_B/(\beta + 1)} \tag{7.147}$$

It is interesting to note that this equation is identical to Eq. (7.143), which governs the operation of the traditional bias circuit, except that V_{CC} replaces V_{BB} and R_C replaces R_E. It follows that to obtain a value of I_E that is insensitive to variation of β, we need to select $R_B/(\beta + 1) \ll R_C$. Note, however, that the value of R_B determines the allowable negative

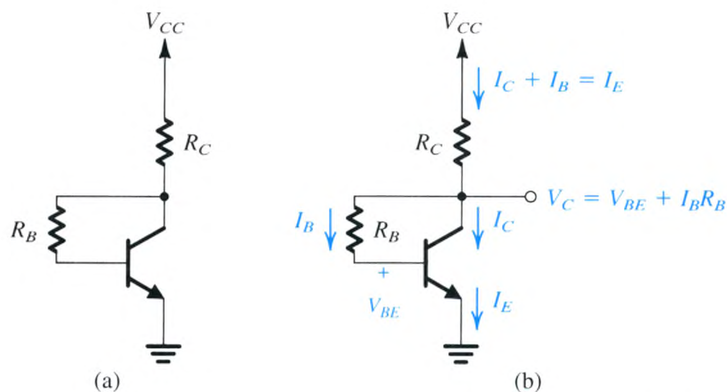


Figure 7.56 (**a**) A common-emitter transistor amplifier biased by a feedback resistor R_B. (**b**) Analysis of the circuit in (**a**).

signal swing at the collector since

$$V_{CB} = I_B R_B = I_E \frac{R_B}{\beta + 1} \tag{7.148}$$

EXERCISE

D7.37 Design the circuit of Fig. 7.56 to obtain a dc emitter current of 1 mA, maximum gain, and a ±2-V signal swing at the collector; that is, design for $V_{CE} = +2.3$ V. Let $V_{CC} = 10$ V and $\beta = 100$.
Ans. $R_B = 162$ kΩ; $R_C = 7.7$ kΩ. Note that if standard 5% resistor values are used (Appendix J), we select $R_B = 160$ kΩ and $R_C = 7.5$ kΩ. This results in $I_E = 1.02$ mA and $V_C = +2.3$ V.

7.5 Discrete-Circuit Amplifiers

With our study of transistor amplifier basics complete, we will now put everything together by presenting practical circuits for discrete-circuit amplifiers. These circuits, which use the amplifier configurations studied in Section 7.3 and the biasing methods of Section 7.4, can be assembled using off-the-shelf discrete transistors, resistors, and capacitors. Though practical and carefully selected to illustrate some important points, the circuits in this section are only *examples* of discrete-circuit transistor amplifiers. Indeed, there is a great variety of such circuits, a number of which are explored in the end-of-chapter problems.

As we mentioned earlier, the vast majority of discrete-circuit amplifiers use BJTs. This is why all but one of the circuits presented in this section use BJTs. Of course, if we wanted, we could use MOSFETs in the same amplifier configurations presented here. Also, the MOSFET is the device of choice in the design of integrated-circuit (IC) amplifiers. We begin our study of IC amplifiers in Chapter 8.

As we will see shortly, the circuits in this section use large capacitors (in the μF range) to couple the signal source to the input of the amplifier, and to couple the amplifier output signal to a load resistance or to the input of another amplifier stage. As well, a large capacitor is used to establish a signal ground at the desired terminal of the transistor (e.g., at the emitter of a CE amplifier). The use of capacitors for these purposes simplifies the design considerably: Since capacitors block dc, we can first carry out the dc bias design and then connect the signal source and load to the amplifier without disturbing the dc design. These amplifiers are therefore known as **capacitively coupled amplifiers.**

7.5.1 A Common-Source (CS) Amplifier

As mentioned in Section 7.3, the common-source (CS) configuration is the most widely used of all MOSFET amplifier circuits. A common-source amplifier using the bias circuit of Fig. 7.50(c) is shown in Fig. 7.57(a). Notice that to establish a **signal ground** (sometimes called an **ac ground**) at the source, we have connected a large capacitor, C_S, between

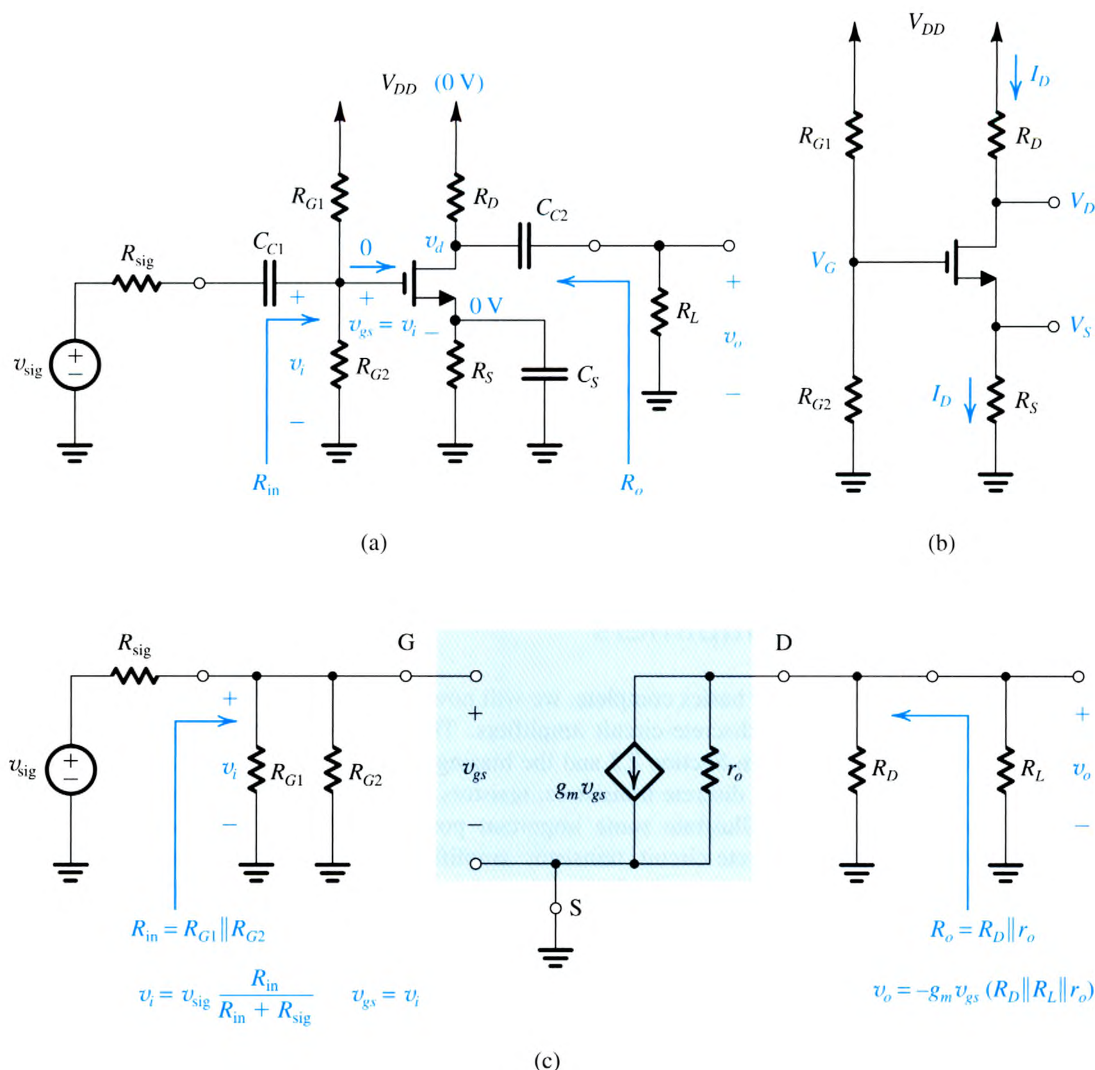


Figure 7.57 (**a**) A common-source amplifier using the classical biasing arrangement of Fig. 7.50(c). Signal quantities are shown in color. (**b**) Circuit for determining the bias point. (**c**) Equivalent circuit and analysis.

the source and ground. This capacitor, usually in the microfarad range, is required to provide a very small impedance (ideally, zero impedance—in effect, a short circuit) at all signal frequencies of interest. In this way, the signal current passes through C_S to ground and thus *bypasses* the resistance R_S; hence, C_S is called a **bypass capacitor**. Obviously, the lower the signal frequency, the less effective the bypass capacitor becomes. We will study this issue in Chapter 10. For our purposes here we will assume that C_S is acting as a perfect short circuit and thus is establishing a zero signal voltage at the MOSFET source.

To prevent disturbances to the dc bias current and voltages, the signal to be amplified, shown as voltage source v_{sig} with an internal resistance R_{sig}, is connected to the gate through a large capacitor C_{C1}. Known as a **coupling capacitor**, C_{C1} is required to act as a perfect

short circuit at all signal frequencies of interest while blocking dc. Here again, we note that as the signal frequency is lowered, the impedance of C_{C1} (i.e., $1/j\omega C_{C1}$) will increase and its effectiveness as a coupling capacitor will be correspondingly reduced. This problem, too, will be considered in Chapter 10 in connection with the dependence of the amplifier operation on frequency. For our purposes here we will assume that C_{C1} is acting as a perfect short circuit as far as the signal is concerned.

The voltage signal resulting at the drain is coupled to the load resistance R_L via another coupling capacitor C_{C2}. We will assume that C_{C2} also acts as a perfect short circuit at all signal frequencies of interest and thus that the output voltage $v_o = v_d$. Note that R_L can be either an actual load resistor, to which the amplifier provides its output voltage signal, or the input resistance of another amplifier stage in cases where more than one stage of amplification is needed. (We will study multistage amplifiers in Chapter 9.)

Since a capacitor behaves as an open circuit at dc, the circuit for performing the dc bias design and analysis is obtained by open-circuiting all capacitors. The resulting circuit is shown in Fig. 7.57(b) and can be designed as discussed in Section 7.4.1.

To determine the characteristics of the CS amplifier of Fig. 7.57(a)—that is, its input resistance, voltage gain, and output resistance—we replace the MOSFET with its hybrid-π small-signal model, replace V_{DD} with a signal ground, and replace all coupling and bypass capacitors with short circuits. The result is the circuit in Fig. 7.57(c). Analysis is straightforward and is shown on the figure, thus

$$R_{\text{in}} = R_{G1} \,\|\, R_{G2} \tag{7.149}$$

which shows that to keep R_{in} high, large values should be used for R_{G1} and R_{G2}, usually in the megohm range. The overall voltage gain G_v is

$$G_v = -\frac{R_{\text{in}}}{R_{\text{in}} + R_{\text{sig}}}\, g_m(R_D \,\|\, R_L \,\|\, r_o) \tag{7.150}$$

Notice that we have taken r_o into account, simply because it is easy to do so. Its effect, however, is usually small (this is not the case for IC amplifiers, as we will see in Chapter 8). Finally, to encourage you to do the small-signal analysis directly on the original circuit diagram, with the MOSFET model used implicitly, we show some of the analysis on the circuit of Fig. 7.57(a).

EXERCISES

D7.38 Design the bias circuit in Fig.7.57(b) for the CS amplifier of Fig. 7.57(a). Assume the MOSFET is specified to have $V_t = 1$ V, $k_n = 4$ mA/V^2, and $V_A = 100$ V. Neglecting the Early effect, design for $I_D = 0.5$ mA, $V_S = 3.5$ V, and $V_D = 6$ V using a power-supply $V_{DD} = 15$ V. Specify the values of R_S and R_D. If a current of 2 μA is used in the voltage divider, specify the values of R_{G1} and R_{G2}. Give the values of the MOSFET parameters g_m and r_o at the bias point.
Ans. $R_S = 7$ kΩ; $R_D = 18$ kΩ; $R_{G1} = 5$ MΩ; $R_{G2} = 2.5$ MΩ; $g_m = 2$ mA/V; $r_o = 200$ kΩ

7.39 For the CS amplifier of Fig. 7.57(a), use the design obtained in Exercise 7.38 to find R_{in}, R_o, and the overall voltage gain G_v when $R_{\text{sig}} = 100$ kΩ and $R_L = 20$ kΩ.
Ans. 1.67 MΩ; 16.5 kΩ; −17.1 V/V

D7.40 As we discussed in Section 7.3, there can be beneficial effects of having an unbypassed resistance R_s in the source lead of the CS amplifier. We can do this in the circuit of Fig. 7.57(a) by splitting the resistance R_S into two resistances: R_s, which is left unbypassed, and $(R_S - R_s)$, across which the bypass capacitor C_S is connected. Now, if, in order to improve linearity of the amplifier in Exercises 7.38 and 7.39, v_{gs} is to be reduced to half its value, what value should R_s have? What would the amplifier gain G_v become? Recall that when R_s is included it becomes difficult to include r_o in the analysis, so neglect it.

Ans. $R_s = 500\ \Omega$; $G_v = -8.9$ V/V

7.5.2 A Common-Emitter Amplifier

The common-emitter (CE) amplifier is the most widely used of all BJT amplifier configurations. Figure 7.58(a) shows a CE amplifier using the classical biasing arrangement of Fig. 7.54(a), the design of which was considered in Section 7.4. The CE circuit in Fig. 7.58(a) is the BJT counterpart of the CS amplifier of Fig. 7.57(a). It utilizes coupling capacitors C_{C1} and C_{C2} and bypass capacitor C_E. Here we assume that these capacitors, while blocking dc, behave as perfect short circuits at all signal frequencies of interest.

To determine the characteristic parameters of the CE amplifier, we replace the BJT with its hybrid-π model, replace V_{CC} with a short circuit to ground, and replace the coupling and bypass capacitors with short circuits. The resulting small-signal equivalent circuit of the CE amplifier is shown in Fig. 7.58(b). The analysis is straightforward and is given in the figure, thus

$$R_{\text{in}} = R_{B1} \parallel R_{B2} \parallel r_\pi \tag{7.151}$$

which tells us that to keep R_{in} relatively high, we should select larger values for R_{B1} and R_{B2} (typically in the range of tens or hundreds of kilohms). This requirement conflicts with the need to keep R_{B1} and R_{B2} low so as to minimize the dependence of the dc current I_C on the transistor β. We discussed this design trade-off in Section 7.4.

The voltage gain G_v is given by

$$G_v = -\frac{R_{\text{in}}}{R_{\text{in}} + R_{\text{sig}}}\, g_m (R_C \parallel R_L \parallel r_o) \tag{7.152}$$

Note that we have taken r_o into account because it is easy to do so. Remember, though, that the effect of this parameter on discrete-circuit amplifier performance is usually small.

EXERCISES

D7.41 Design the bias circuit of the CE amplifier of Fig. 7.58(a) to obtain $I_E = 0.5$ mA and $V_C = +6$ V. Design for a dc voltage at the base of 5 V and a current through R_{B2} of 50 μA. Let $V_{CC} = +15$ V, $\beta = 100$, and $V_{BE} \simeq 0.7$ V. Specify the values of R_{B1}, R_{B2}, R_E, and R_C. Also give the

values of the BJT small-signal parameters g_m, r_π, and r_o at the bias point. (For the calculation of r_o, let $V_A = 100$ V.)

Ans. $R_{B1} = 182\text{ k}\Omega$; $R_{B2} = 100\text{ k}\Omega$; $R_E = 8.6\text{ k}\Omega$; $R_C = 18\text{ k}\Omega$; $g_m = 20$ mA/V, $r_\pi = 5\text{ k}\Omega$, $r_o = 200\text{ k}\Omega$

7.42 For the amplifier designed in Exercise 7.41, find R_{in}, R_o, and G_v when $R_{sig} = 10\text{ k}\Omega$ and $R_L = 20\text{ k}\Omega$.

Ans. $R_{in} = 4.64\text{ k}\Omega$; $R_o = 16.51\text{ k}\Omega$; $G_v = -57.3$ V/V

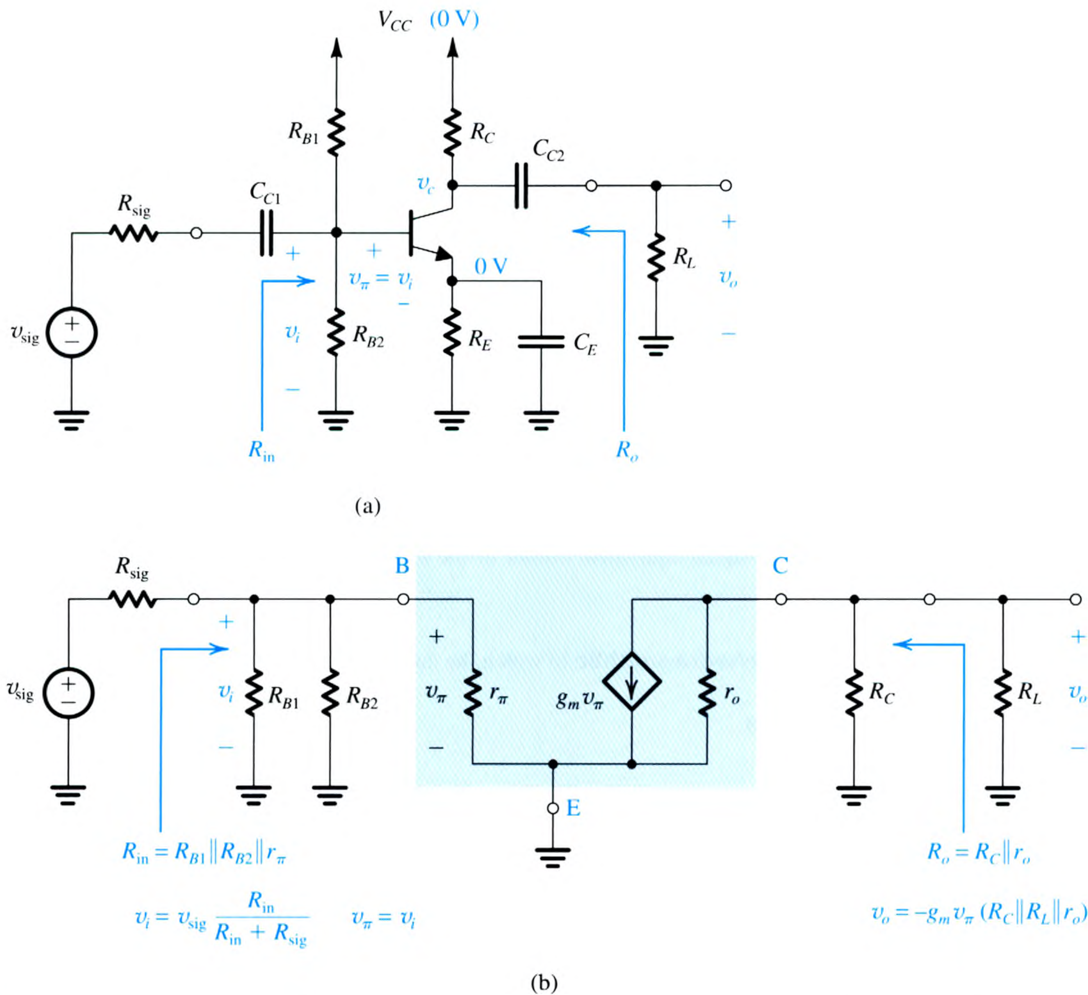


Figure 7.58 (**a**) A common-emitter amplifier using the classical biasing arrangement of Fig. 7.54(a). Signal quantities are shown in color. (**b**) Equivalent circuit and analysis.

Video Example VE 7.3 Design and Analysis of a Common-Emitter Amplifier

In the circuit of Fig. VE7.3, v_{sig} is a small sine-wave signal with zero average. The transistor β is 100.

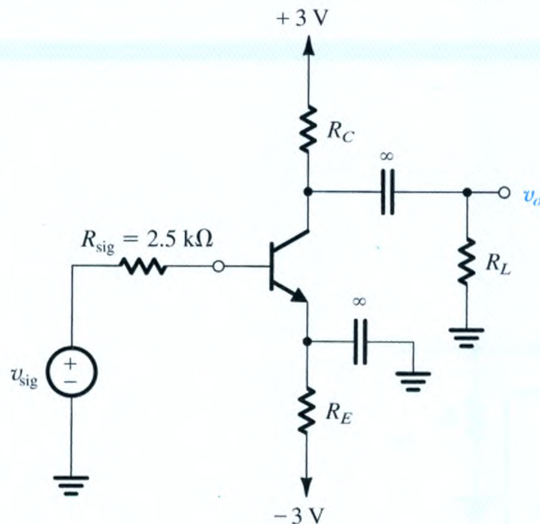


Figure VE7.3 Circuit for Video Example 7.3.

(a) Find the value of R_E to establish a dc emitter current of about 0.5 mA.
(b) Find R_C to establish a dc collector voltage of about +0.5 V.
(c) For $R_L = 10\ \text{k}\Omega$, draw the small-signal equivalent circuit of the amplifier and determine its overall voltage gain.


VE 7.3

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 7.129

7.5.3 A Common-Emitter Amplifier with an Emitter Resistance R_e

As we discussed in Section 7.3.4, it helps to include a small resistance in the transistor emitter lead. We can do this in the circuit of Fig. 7.58(a) by splitting the emitter bias resistance R_E into two components: an unbypassed resistance R_e, and a resistance $(R_E - R_e)$ across which the bypass capacitor C_E is connected. The resulting circuit is shown in Fig. 7.59(a), and its small-signal model is shown in Fig. 7.59(b). In the latter we use the T model of the BJT because it results in much simpler analysis (recall that this is always

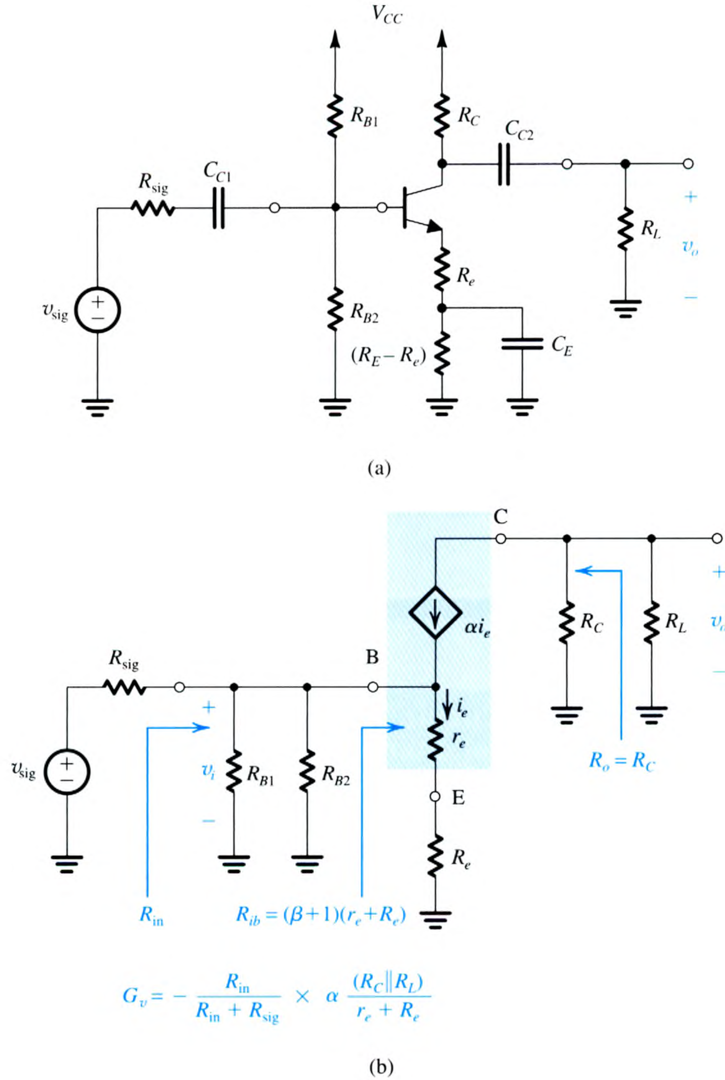


Figure 7.59 (**a**) A common-emitter amplifier with an unbypassed emitter resistance R_e. (**b**) The amplifier small-signal model and analysis.

the case when a resistance is connected in series with the emitter). Also note that we have not included r_o, because it would complicate the analysis significantly, which is not worth the trouble, given that r_o has little effect on the performance of discrete-circuit amplifiers.

Analysis of the circuit in Fig. 7.59(b) is straightforward and is shown in the figure. Thus,

$$R_{in} = R_{B1} \| R_{B2} \| (\beta+1)(r_e + R_e)$$

$$= R_{B1} \| R_{B2} \| [r_\pi + (\beta+1)R_e] \tag{7.153}$$

Here we see that including R_e increases R_{in} because it increases the input resistance looking into the base by adding a component $(\beta + 1)R_e$ to r_π. The overall voltage gain G_v is

$$G_v = -\frac{R_{in}}{R_{in}+R_{sig}} \times \alpha \frac{\text{Total resistance in collector}}{\text{Total resistance in emitter}}$$

$$= -\alpha \frac{R_{in}}{R_{in}+R_{sig}} \frac{R_C \| R_L}{r_e + R_e} \qquad (7.154)$$

EXERCISE

7.43 For the amplifier designed in Exercise 7.41 and analyzed in Exercise 7.42, we want to raise R_{in} to 10 kΩ. What is the required value of R_e, and what does the overall voltage gain G_v become?
Ans. $R_e = 67.2\ \Omega$; $G_v = -40$ V/V

Video Example VE 7.4 Discrete-Circuit CE Amplifier with Emitter-Degeneration

In the circuit of Fig. VE7.4, the BJT is biased with a constant-current source, and v_{sig} is a small sine-wave signal. Find R_{in} and the gain v_o/v_{sig}. Assume $\beta = 100$. If the amplitude of the signal v_{be} is to be limited to 5 mV, what is the largest allowable signal at the input? What is the corresponding signal at the output?

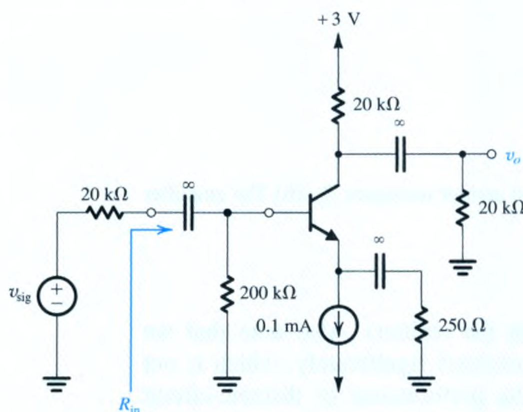


Figure VE7.4 Circuit for Video Example 7.4.



VE 7.4

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 7.131

7.5.4 A Common-Base (CB) Amplifier

Figure 7.60(a) shows a CB amplifier using the biasing arrangement of Fig. 7.55. Note that having two power supplies, V_{CC} and $-V_{EE}$, enables us to connect the base directly to the ground, obviating the need for a large bypass capacitor to establish a signal ground at the base.

The small-signal equivalent circuit of the CB amplifier is shown in Fig 7.60(b). As expected, we have used the T model of the BJT and have not included r_o. Including r_o would complicate the analysis significantly without making much difference to the results. From the circuit in Fig. 7.60(b) we find

$$R_{in} = r_e \| R_E \simeq r_e \simeq 1/g_m$$

which as expected can be very small, causing v_i to be a small fraction of v_{sig},

$$v_i = v_{sig} \frac{R_{in}}{R_{in} + R_{sig}}$$

Now,

$$i_e = -\frac{v_i}{r_e}$$

and

$$v_o = -\alpha i_e (R_C \| R_L)$$

Thus, the overall voltage gain is given by

$$G_v = \alpha \frac{R_{in}}{R_{in} + R_{sig}} \frac{R_C \| R_L}{r_e} = \frac{R_{in}}{R_{in} + R_{sig}} g_m (R_C \| R_L) \tag{7.155}$$

EXERCISE

D7.44 Design the CB amplifier of Fig. 7.60(a) to provide an input resistance R_{in} that matches the source resistance of a cable with a characteristic resistance of 50 Ω. Assume that $R_E \gg r_e$. The available power supplies are ±5 V and $R_L = 8$ kΩ. Design for a dc collector voltage $V_C = +1$ V. Specify the values of R_C and R_E. What is the overall voltage gain? If v_{sig} is a sine wave with a peak amplitude of 10 mV, what is the peak amplitude of the output voltage? Let $\alpha \simeq 1$.
Ans. $R_C = 8$ kΩ; $R_E = 8.6$ kΩ; 40 V/V; 0.4 V

7.5.5 An Emitter Follower

Figure 7.61(a) shows an emitter follower designed using the bias arrangement of Fig. 7.55 with two power supplies, V_{CC} and $-V_{EE}$. The bias resistance R_B affects the input resistance of the follower and should be chosen as large as possible while limiting the dc voltage drop

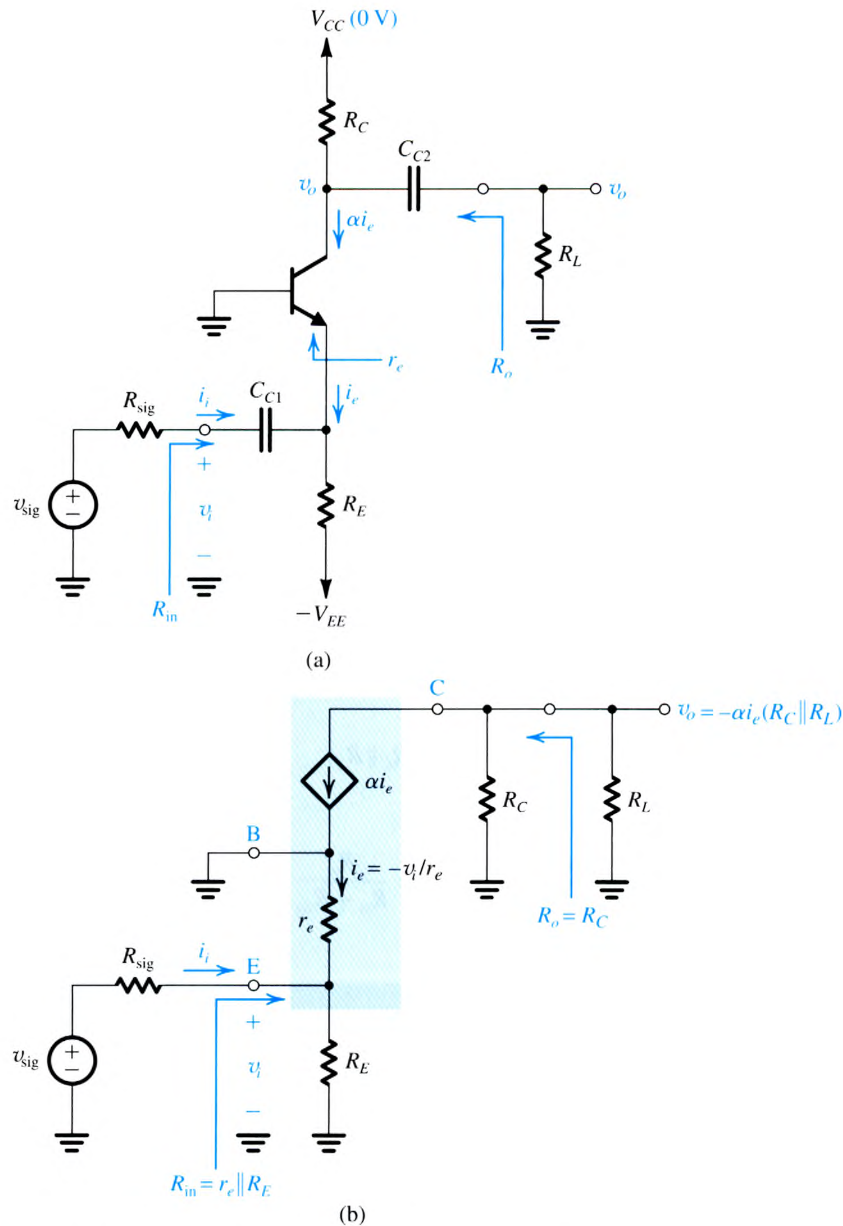


Figure 7.60 (**a**) A common-base amplifier using the structure of Fig. 7.55 with R_B omitted (since the base is grounded). Small-signal quantities are shown in color. (**b**) Equivalent circuit obtained by replacing the transistor with its T model.

across it to a small fraction of V_{EE}; otherwise the dependence of the bias current I_C on β can become unacceptably large.

Figure 7.61(b) shows the small-signal equivalent circuit of the emitter follower. Here, as expected, we have replaced the BJT with its T model and included r_o (since this can be done

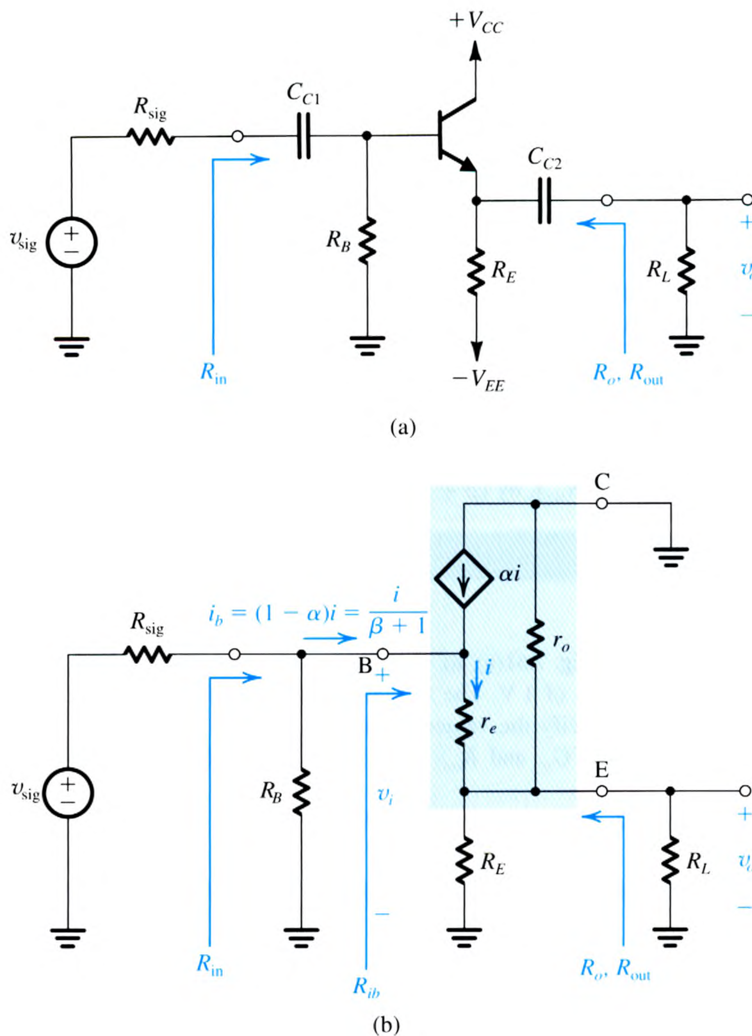


Figure 7.61 (**a**) An emitter-follower circuit. (**b**) Small-signal equivalent circuit of the emitter follower with the transistor replaced by its T model. Note that r_o is included because it is easy to do so. Normally, its effect on performance is small.

very simply). We can see that the input resistance of the emitter follower is

$$R_{in} = R_B \| R_{ib} \tag{7.156}$$

where R_{ib}, the input resistance looking into the base, can be obtained by using the resistance-reflection rule. Toward that end, note that r_o appears in parallel with R_E and R_L (which is why it can be easily taken into account). Thus,

$$R_{ib} = (\beta + 1)[r_e + (R_E \| r_o \| R_L)] \tag{7.157}$$

We can find the overall voltage gain by tracking the signal transmission from source to load,

$$v_i = v_{sig} \frac{R_{in}}{R_{in} + R_{sig}} \tag{7.158}$$

and v_o is obtained as the fraction of v_i determined by the voltage-divider ratio,

$$v_o = v_i \frac{R_E \| r_o \| R_L}{r_e + (R_E \| r_o \| R_L)} \tag{7.159}$$

Thus,

$$G_v \equiv \frac{v_o}{v_{sig}} = \frac{R_{in}}{R_{in} + R_{sig}} \frac{(R_E \| r_o \| R_L)}{r_e + (R_E \| r_o \| R_L)} \tag{7.160}$$

Finally, we can find the output resistance R_{out} by short-circuiting v_{sig} and looking back into the output terminal, excluding R_L, as

$$R_{out} = r_o \| R_E \| \left[r_e + \frac{R_B \| R_{sig}}{\beta + 1} \right] \tag{7.161}$$

where we have used the inverse resistance-reflection rule, namely, dividing the total resistance in the base, $(R_B \| R_{sig})$, by $(\beta + 1)$.

EXERCISE

D7.45 Design the emitter follower of Fig. 7.61(a) to operate at a dc emitter current $I_E = 1$ mA. Allow a dc voltage drop across R_B of 1 V. The available power supplies are ±5 V, $\beta = 100$, $V_{BE} = 0.7$ V, and $V_A = 100$ V. Specify the values required for R_B and R_E. If $R_{sig} = 50$ kΩ and $R_L = 1$ kΩ, find R_{in}, v_i/v_{sig}, v_o/v_i, G_v, and R_{out}. (*Note*: In the bias design, neglect the Early effect.)

Ans. $R_B = 100$ kΩ; $R_E = 3.3$ kΩ; 44.3 kΩ; 0.469 V/V; 0.968 V/V; 0.454 V/V; 320 Ω

7.5.6 The Amplifier Frequency Response

So far, we have assumed that the gain of transistor amplifiers is constant independent of the frequency of the input signal. This implies that transistor amplifiers have infinite bandwidth, which of course is not true. To illustrate, we show in Fig. 7.62 a sketch of the gain of a CS or CE amplifier such as those shown in Figs. 7.57 and 7.58, respectively, versus frequency. Observe that, indeed, the gain remains almost constant over a wide frequency range. This is the useful frequency range of operation for the particular amplifier. So far, we have been assuming that our amplifiers are operating in this frequency band, called the **midband**.

Figure 7.62 indicates that at lower frequencies, the magnitude of the amplifier gain falls off. This is because the coupling and bypass capacitors no longer have low impedances. Recall that we assumed that their impedances were small enough to act as short circuits. Although this can be true at midband frequencies, as the frequency of the input signal is lowered, the reactance $1/j\omega C$ of each of these capacitors becomes significant, and it can be shown that, as a result, the overall voltage gain of the amplifier decreases.

Figure 7.62 also indicates that the gain of the amplifier falls off at the high-frequency end. This is due to the internal capacitive effects in the BJT and the MOSFET. In Chapter 10 we will study the internal capacitive effects of both transistor types and will augment their hybrid-π models with capacitances that model these effects.

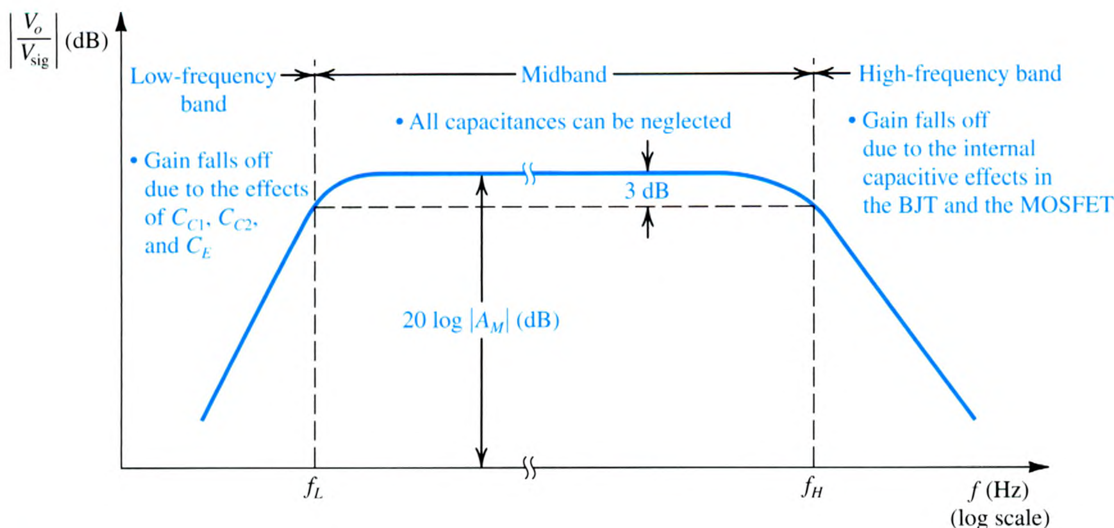


Figure 7.62 Sketch of the magnitude of the gain of a CS (Fig. 7.57) or CE (Fig. 7.58) amplifier versus frequency. The graph delineates the three frequency bands relevant to frequency-response determination.

We will also undertake a detailed study of the frequency response of transistor amplifiers in Chapter 10. For the time being, it is important to realize that for every transistor amplifier, there is a finite band over which the gain is almost constant. The boundaries of this useful frequency band, or midband, are the two frequencies f_L and f_H at which the gain drops by a certain number of decibels (usually 3 dB) below its value at midband. As shown in Fig. 7.62, the amplifier **bandwidth**, or 3-dB bandwidth, is defined as the difference between the lower (f_L) and upper or higher (f_H) 3-dB frequencies:

$$BW = f_H - f_L$$

and since usually $f_L \ll f_H$,

$$BW \simeq f_H$$

A figure of merit for the amplifier is its **gain–bandwidth product**, defined as

$$GB = |A_M|BW$$

where $|A_M|$ is the magnitude of the amplifier gain in the midband. You will see in Chapters 10 and 11 that in amplifier design it is usually possible to trade off gain for bandwidth—for instance, by including resistance R_e in the emitter of the CE amplifier.

Summary

- The essence of the MOSFET (the BJT) as an amplifier is that when the transistor is operated in the active region, v_{GS} controls i_D (v_{BE} controls i_C) in the manner of a voltage-controlled current source. When the device is dc biased in the active region, and the signal v_{gs} (v_{be}) is kept small, the operation becomes almost linear, with $i_d = g_m v_{gs}$ ($i_c = g_m v_{be}$).

- The most fundamental parameter in characterizing the small-signal linear operation of a transistor is the transconductance g_m. For a MOSFET, $g_m = \mu_n C_{ox}(W/L)V_{OV} = \sqrt{2\mu_n C_{ox}(W/L)I_D} = 2I_D/V_{OV}$; and for the BJT, $g_m = I_C/V_T$.

- A systematic procedure for analyzing a transistor amplifier circuit is presented in Table 7.1. Tables 7.2 and 7.3 present the small-signal models for the MOSFET and the BJT, respectively. These tables can be found in the Summary Tables supplement online at www.oup.com/he/sedra-smith8e.

- When a resistance is connected in series with the source (or emitter), the T model is the most convenient to use.

- The three basic configurations of MOS and BJT amplifiers are presented in Fig. 7.35. Their characteristic parameter values are provided in Table 7.4 (for the MOS case) and in Table 7.5 (for the BJT case). Both tables are in the Summary Tables supplement online at www.oup.com/he/sedra-smith8e.

- The CS amplifier, has (ideally) infinite input resistance and a reasonably high gain but a rather high output resistance and a limited high-frequency response (more on the latter point in Chapter 10). It is used to obtain most of the gain in a cascade amplifier. Similar remarks apply to the CE amplifier, except that it has a relatively low input resistance ($r_\pi = \beta/g_m$) arising from the finite base current of the BJT (finite β). Its voltage gain, however, can be larger than that of the CS amplifier because of the higher values of g_m obtained with BJTs.

- Adding a resistance R_s in the source of a CS amplifier (a resistance R_e in the emitter of a CE amplifier) can benefit the design by raising the input resistance of the CE amplifier, increasing linearity, and extending the useful amplifier bandwidth, at the expense of reducing the gain, all by a factor equal to $(1+g_mR_s)$ [$(1+g_mR_e)$ for the BJT case].

- The CG (CB) amplifier has a low input resistance and thus, used alone, it has limited and specialized applications. However, its excellent high-frequency response makes it attractive in combination with the CS (CE) amplifier (Chapters 8 and 10).

- The source follower has (ideally) infinite input resistance, a voltage gain lower than but close to unity, and a low output resistance. It is used as a voltage buffer and as the output stage of a multistage amplifier. Similar remarks apply to the emitter follower except that its input resistance, though large, is finite. Specifically, the emitter follower multiplies the total resistance in the emitter by $(\beta+1)$ before presenting it to the signal source.

- The resistance-reflection rule is useful in the analysis of BJT amplifier circuits: All resistances in the emitter circuit including the emitter resistance r_e can be reflected to the base side by multiplying them by $(\beta+1)$. Conversely, we can reflect all resistances in the base circuit to the emitter side by dividing them by $(\beta+1)$.

- In the analysis and design of discrete-circuit amplifiers, it is rarely necessary to take the transistor output resistance r_o into account. In some situations, however, r_o can be easily taken into account; specifically in the CS (CE) amplifier and in the source (emitter) follower. (In IC amplifiers, r_o must always be taken into account.)

- A key step in the design of transistor amplifiers is to bias the transistor to operate at an appropriate point in the active region. A good bias design ensures that the parameters of the operating point (I_D, V_{OV}, and V_{DS} for the MOSFET; I_C and V_{CE} for the BJT) are predictable and stable and do not vary by large amounts when the transistor is replaced by another of the same type. The bias methods studied in this chapter are suited for discrete-circuit amplifiers only because they use large coupling and bypass capacitors.

- Discrete-circuit amplifiers predominantly use BJTs. The opposite is true for IC amplifiers, where the device of choice is the MOSFET.

PROBLEMS

Computer Simulation Problems

SIM Problems identified by this icon are intended to demonstrate the value of using SPICE simulation to verify hand analysis and design, and to investigate important issues such as allowable signal swing and amplifier nonlinear

---

problems with blue numbers are considered essential; * = difficult problem; ** = more difficult; *** = very challenging
SIM = simulation; **D** = design problem; ● = see related video example

distortion. Instructions to assist in setting up simulations for all the indicated problems can be found in the corresponding files on the website. Note that if a particular parameter value is not specified in the problem statement, you are to make a reasonable assumption.

Section 7.1: Basic Principles

7.1 For the MOS amplifier of Fig. 7.2(a) with $V_{DD} = 3$ V, $V_t = 0.5$ V, $k_n = 10$ mA/V^2, and $R_D = 15$ kΩ, determine the coordinates of the active-region segment (AB) of the VTC [Fig. 7.2(b)].

D 7.2 For the MOS amplifier of Fig. 7.2(a) with $V_{DD} = 3$ V and $k_n = 5$ mA/V^2, we want to have the end point of the VTC, point B, at $V_{DS} = 0.3$ V. What value of R_D do we need? If the transistor is replaced with another having twice the value of the transconductance parameter k_n, what new value of R_D do we need?

D 7.3 We want to bias the MOS amplifier of Fig. 7.3 at point Q for which $V_{OV} = 0.2$ V and $V_{DS} = 1$ V. Find the required value of R_D when $V_{DD} = 3$ V, $V_t = 0.5$ V, and $k_n = 10$ mA/V^2. Also specify the coordinates of the VTC end point B. What is the small-signal voltage gain of this amplifier? Assuming linear operation, what is the maximum allowable negative signal swing at the output? What is the corresponding peak input signal?

7.4 The MOS amplifier of Fig. 7.4(a), when operated with $V_{DD} = 2$ V, has a maximum small-signal voltage gain magnitude of 14 V/V. Find V_{OV} and V_{DS} for bias point Q at which a voltage gain of -12 V/V is obtained.

7.5 Consider the amplifier of Fig. 7.4(a) for the case $V_{DD} = 5$ V, $R_D = 24$ kΩ, $k_n'(W/L) = 1$ mA/V^2, and $V_t = 1$ V.

(a) Find the coordinates of the two end points of the saturation-region segment of the amplifier transfer characteristic, that is, points A and B in Fig. 7.4(b).

(b) If the amplifier is biased to operate with an overdrive voltage V_{OV} of 0.5 V, find the coordinates of the bias point Q on the transfer characteristic. Also, find the value of I_D and of the incremental gain A_v at the bias point.

(c) For the situation in (b), and disregarding the distortion caused by the MOSFET's square-law characteristic, what is the largest amplitude of a sine-wave voltage signal that can be applied at the input while the transistor remains in saturation? What is the amplitude of the output voltage signal that results? What gain value does the combination of these amplitudes imply? By what percentage is this gain value different from the incremental gain value calculated above? Why is there a difference?

7.6 Various measurements are made on an NMOS amplifier for which the drain resistor R_D is 20 kΩ. First, dc measurements show the voltage across the drain resistor, V_{RD}, to be 2 V and the gate-to-source bias voltage to be 0.75 V. Then, ac measurements with small signals show the voltage gain to be –16 V/V. What is the value of V_t for this transistor? If the process transconductance parameter k_n' is 400 μA/V^2, what is the MOSFET's W/L?

***7.7** The expression for the incremental voltage gain A_v given in Eq. (7.16) can be written as

$$A_v = -\frac{2(V_{DD} - V_{DS})}{V_{OV}}$$

where V_{DS} is the bias voltage at the drain. This indicates that for given values of V_{DD} and V_{OV}, the gain magnitude can be increased by biasing the transistor at a lower V_{DS}. This, however, reduces the allowable output signal swing in the negative direction. Assuming linear operation around the bias point, show that the largest possible negative output signal peak $\hat{v}_o$ achievable while the transistor remains saturated is

$$\hat{v}_o = (V_{DS} - V_{OV}) \Big/ \left(1 + \frac{1}{|A_v|}\right)$$

For $V_{DD} = 5$ V and $V_{OV} = 0.5$ V, provide a table of values for A_v, $\hat{v}_o$, and the corresponding $\hat{v}_i$ for $V_{DS} = 1$ V, 1.5 V, 2 V, and 2.5 V. If $k_n'(W/L) = 1$ mA/V^2, find I_D and R_D for the design for which $V_{DS} = 1$ V.

D *7.8 Design the MOS amplifier of Fig. 7.4(a) to obtain maximum gain while allowing for an output voltage swing of at least ±0.2 V. Let $V_{DD} = 2$ V, and use an overdrive voltage of approximately 0.2 V.

(a) Specify V_{DS} at the bias point.

(b) What is the gain achieved? What is the signal amplitude $\hat{v}_{gs}$ that results in the 0.2-V signal amplitude at the output?

(c) If the dc bias current in the drain is to be 100 μA, what value of R_D is needed?

(d) If $k_n' = 400$ μA/V^2, what W/L ratio is required for the MOSFET?

***7.9** Figure P7.9 shows an amplifier in which the load resistor R_D has been replaced with another NMOS transistor Q_2 connected as a two-terminal device. Note that because v_{DG} of Q_2 is zero, it will be operating in saturation at all

times, even when $v_I = 0$ and $i_{D2} = i_{D1} = 0$. Note also that the two transistors conduct equal drain currents. Using $i_{D1} = i_{D2}$, show that for the range of v_I over which Q_1 is operating in saturation, that is, for

$$V_{t1} \le v_I \le v_O + V_{t1}$$

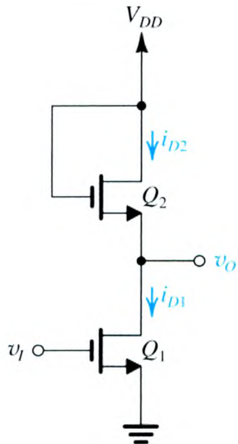


Figure P7.9

the output voltage will be given by

$$v_O = V_{DD} - V_t + \sqrt{\frac{(W/L)_1}{(W/L)_2}}V_t - \sqrt{\frac{(W/L)_1}{(W/L)_2}}v_I$$

where we have assumed $V_{t1} = V_{t2} = V_t$. Thus the circuit functions as a linear amplifier, even for large input signals. For $(W/L)_1 = (50\ \mu\text{m}/0.5\ \mu\text{m})$ and $(W/L)_2 = (5\ \mu\text{m}/0.5\ \mu\text{m})$, find the voltage gain.

7.10 A BJT amplifier circuit like the one in Fig. 7.6 is operated with $V_{CC} = +3$ V and is biased at $V_{CE} = +0.5$ V. Find the voltage gain, the maximum allowed output negative swing without the transistor entering saturation, and the corresponding maximum input signal permitted.

7.11 For the amplifier circuit in Fig. 7.6 with $V_{CC} = +5$ V and $R_C = 1$ kΩ, find V_{CE} and the voltage gain at the following dc collector bias currents: 0.5 mA, 1 mA, 2.5 mA, 4 mA, and 4.5 mA. For each, give the maximum possible positive- and negative-output signal swings as determined by the need to keep the transistor in the active region. Present your results in a table.

D 7.12 Consider the CE amplifier circuit of Fig. 7.6 when operated with a dc supply $V_{CC} = +3$ V. You want to find the point at which the transistor should be biased; that is, the value of V_{CE} so that the output sine-wave signal v_{ce} resulting from an input sine-wave signal v_{be} of 5-mV peak amplitude has the maximum possible magnitude. What is the peak amplitude of the output sine wave and the value of the gain obtained? Assume linear operation around the bias point. (*Hint:* To obtain the maximum possible output amplitude for a given input, you need to bias the transistor as close to the edge of saturation as possible without entering saturation at any time, that is, without v_{CE} decreasing below 0.3 V.)

7.13 A designer considers a number of low-voltage BJT amplifier designs utilizing power supplies with voltage V_{CC} of 1.0, 1.5, 2.0, or 3.0 V. For transistors that saturate at $V_{CE} =$ 0.3 V, what is the largest possible voltage gain achievable with each of these supply voltages? If in each case biasing is adjusted so that $V_{CE} = V_{CC}/2$, what gains are achieved? If a negative-going output signal swing of 0.4 V is required, at what V_{CE} should the transistor be biased to obtain maximum gain? What is the gain achieved with each of the supply voltages? (Notice that all of these gains are independent of the value of I_C chosen!)

D *7.14 A BJT amplifier such as that in Fig. 7.6 is to be designed to support relatively undistorted sine-wave output signals of peak amplitudes P volts without the BJT entering saturation or cutoff and to have the largest possible voltage gain, denoted A_v V/V. Show that the minimum supply voltage V_{CC} needed is given by

$$V_{CC} = V_{CE\text{sat}} + P + |A_v|V_T$$

Also, find V_{CC}, specified to the nearest 0.5 V, for the following situations:

(a) $A_v = -20$ V/V, $P = 0.2$ V
(b) $A_v = -50$ V/V, $P = 0.5$ V
(c) $A_v = -100$ V/V, $P = 0.5$ V
(d) $A_v = -100$ V/V, $P = 1.0$ V
(e) $A_v = -200$ V/V, $P = 1.0$ V
(f) $A_v = -500$ V/V, $P = 1.0$ V
(g) $A_v = -500$ V/V, $P = 2.0$ V

7.15 The transistor in the circuit of Fig. P7.15 is biased at a dc collector current of 0.2 mA. What is the voltage gain? (*Hint:* Use Thévenin's theorem to convert the circuit to the form in Fig. 7.6.)

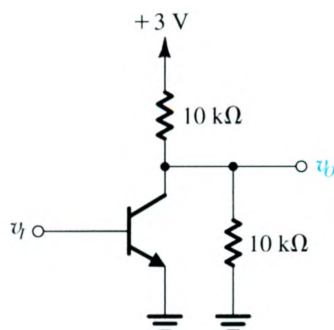


Figure P7.15

7.16 Sketch and label the voltage-transfer characteristic of the *pnp* amplifier shown in Fig. P7.16.

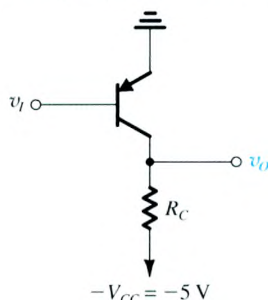


Figure P7.16

***7.17** In deriving the expression for small-signal voltage gain A_v in Eq. (7.21) we neglected the Early effect. Derive this expression including the Early effect by substituting

$$i_C = I_S e^{v_{BE}/V_T}\left(1 + \frac{v_{CE}}{V_A}\right)$$

in Eq. (7.4) and including the factor $(1 + V_{CE}/V_A)$ in Eq. (7.11). Show that the gain expression changes to

$$A_v = \frac{-I_C R_C / V_T}{\left[1 + \frac{I_C R_C}{V_A + V_{CE}}\right]} = -\frac{(V_{CC} - V_{CE})/V_T}{\left[1 + \frac{V_{CC} - V_{CE}}{V_A + V_{CE}}\right]}$$

When $V_{CC} = 5$ V and $V_{CE} = 2$ V, what is the gain without and with the Early effect taken into account? Let $V_A = 50$ V.

7.18 When the amplifier circuit of Fig. 7.6 is biased with a certain V_{BE}, the dc voltage at the collector is found to be +1 V. For $V_{CC} = +3$ V and $R_C = 2$ kΩ, find I_C and the small-signal voltage gain. For a change $\Delta v_{BE} = +5$ mV, calculate the resulting Δv_O. Calculate it two ways: by using the transistor exponential characteristic Δi_C, and approximately, using the small-signal voltage gain. Repeat for $\Delta v_{BE} = -5$ mV. Summarize your results in a table.

***7.19** Consider the amplifier circuit of Fig. 7.6 when operated with a supply voltage $V_{CC} = +3$V.

(a) What is the theoretical maximum voltage gain that this amplifier can provide?
(b) What value of V_{CE} must this amplifier be biased at to provide a voltage gain of –60 V/V?
(c) If the dc collector current I_C at the bias point in (b) is to be 0.5 mA, what value of R_C should be used?
(d) What is the value of V_{BE} required to provide the bias point mentioned above? Assume that the BJT has $I_S = 10^{-15}$ A.
(e) If a sine-wave signal v_{be} with a 5-mV peak amplitude is superimposed on V_{BE}, find the corresponding output voltage signal v_{ce} that will be superimposed on V_{CE} assuming linear operation around the bias point.
(f) Characterize the signal current i_c that will be superimposed on the dc bias current I_C.
(g) What is the value of the dc base current I_B at the bias point? Assume $\beta = 100$. Characterize the signal current i_b that will be superimposed on the base current I_B.
(h) Dividing the amplitude of v_{be} by the amplitude of i_b, evaluate the incremental (or small-signal) input resistance of the amplifier.
(i) Sketch and clearly label correlated graphs for v_{BE}, v_{CE}, i_C, and i_B versus time. Note that each graph consists of a dc or average value and a superimposed sine wave. Be careful of the phase relationships of the sine waves.

7.20 The essence of transistor operation is that a change in v_{BE}, Δv_{BE}, produces a change in i_C, Δi_C. By keeping Δv_{BE} small, Δi_C is approximately linearly related to Δv_{BE}, $\Delta i_C = g_m \Delta v_{BE}$, where g_m is known as the transistor transconductance. By passing Δi_C through R_C, an output voltage signal Δv_O is obtained. Use the expression for the small-signal voltage gain in Eq. (7.20) to derive an expression for g_m. Find the value of g_m for a transistor biased at $I_C = 0.5$ mA.

7.21 The purpose of this problem is to illustrate the application of graphical analysis to the circuit shown in Fig. P7.21. Sketch $i_C - v_{CE}$ characteristic curves for the BJT for $i_B =$ 10 μA, 20 μA, 30 μA, and 40 μA. Assume the lines to be horizontal (i.e., neglect the Early effect), and let $\beta = 100$. For $V_{CC} = 5$ V and $R_C = 1$ kΩ, sketch the load line. What peak-to-peak collector voltage swing will result for i_B varying over the range 10 μA to 40 μA? If the BJT is biased at $V_{CE} = \frac{1}{2}V_{CC}$, find the value of I_C and I_B. If at this current $V_{BE} = 0.7$ V and if $R_B = 100$ kΩ, find the required value of V_{BB}.

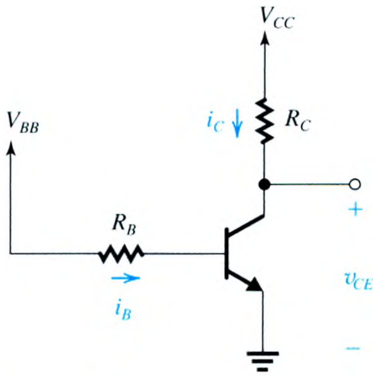


Figure P7.21

****7.22** Sketch the $i_C - v_{CE}$ characteristics of an *npn* transistor having $\beta = 100$ and $V_A = 100$ V. Sketch characteristic curves for $i_B = 20\ \mu$A, 50 μA, 80 μA, and 100 μA. For the purpose of this sketch, assume that $i_C = \beta i_B$ at $v_{CE} = 0$. Also, sketch the load line obtained for $V_{CC} = 10$ V and $R_C = 1$ kΩ. If the dc bias current into the base is 50 μA, write the equation for the corresponding $i_C - v_{CE}$ curve. Also, write the equation for the load line, and solve the two equations to obtain V_{CE} and I_C. If the input signal causes a sinusoidal signal of 30-μA peak amplitude to be superimposed on I_B, find the corresponding signal components of i_C and v_{CE}.

Section 7.2: Small-Signal Operation and Models

7.23 This problem investigates the nonlinear distortion introduced by a MOSFET amplifier. Let the signal v_{gs} be a sine wave with amplitude V_{gs}, and substitute $v_{gs} = V_{gs} \sin \omega t$ in Eq. (7.28). Using the trigonometric identity $\sin^2\theta = \frac{1}{2} - \frac{1}{2}\cos 2\theta$, show that the ratio of the signal at frequency 2ω to that at frequency ω, expressed as a percentage (known as the second-harmonic distortion) is

$$\text{Second-harmonic distortion} = \frac{1}{4}\frac{V_{gs}}{V_{OV}} \times 100$$

If in a particular application V_{gs} is 10 mV, find the minimum overdrive voltage at which the transistor should be operated so that the second-harmonic distortion is kept to less than 1%.

7.24 For the amplifier in Fig. 7.10, let $V_{DD} = 2$ V, $R_D = 15$ kΩ, $V_t = 0.5$ V, $k_n' = 0.4$ mA/V^2, $W/L = 12.5$, $V_{GS} = 0.7$ V, and $\lambda = 0$.

(a) Find the dc current I_D and the dc voltage V_{DS}.
(b) Find g_m.
(c) Find the voltage gain.
(d) If $v_{gs} = 0.015 \sin \omega t$ volts, find v_{ds} assuming that the small-signal approximation holds. What are the minimum and maximum values of v_{DS}?
(e) Use Eq. (7.28) to determine the various components of i_D. Using the identity $\sin^2 \omega t = \frac{1}{2} - \frac{1}{2}\cos 2\omega t$, determine the second-harmonic distortion defined as the ratio of the component of i_D of frequency 2ω to the component of i_D of frequency ω, multiplied by 100.

7.25 Consider the MOSFET amplifier of Fig. 7.10 for the case $V_t = 0.4$ V, $k_n = 10$ mA/V^2, $V_{GS} = 0.6$ V, $V_{DD} = 1.8$ V, and $R_D = 6.8$ kΩ.

(a) Find the dc quantities I_D and V_{DS}.
(b) Calculate the value of g_m at the bias point.
(c) Calculate the value of the voltage gain.
(d) If the MOSFET has $\lambda = 0.2$ V^{-1}, find r_o at the bias point and calculate the voltage gain.

D 7.26 The NMOS amplifier of Fig. 7.10 is to be designed to provide a 0.2-V peak output signal. Assume $V_{DD} = 1.8$ V and $R_D = 10$ kΩ. If the gain is to be −10 V/V, what g_m is required? Bias the amplifier as close to the edge of the saturation region as possible consistent with the required signal swing. Specify the required values of V_{OV} and I_D. If $k_n' = 380\ \mu$A/V^2 what W/L ratio is required? If $V_t = 0.4$ V, find V_{GS}.

7.27 An NMOS transistor has $\mu_n C_{ox} = 400\ \mu$A/V^2, $W/L = 20$, $V_t = 0.5$ V, and $V_A = 5$ V. Find g_m and r_o when (a) the bias voltage $V_{GS} = 0.75$ V, (b) the bias current $I_D = 0.5$ mA.

7.28 In the table on page 479, for MOS transistors operating under a variety of conditions, complete as many entries as possible. Although some data is not available, it is always possible to calculate g_m using either Eq. (7.40), (7.41), or (7.42). Assume $\mu_n = 500$ cm^2/V·s, $\mu_p = 250$ cm^2/V·s, and $C_{ox} = 0.4$ fF/μm^2.

7.29 An NMOS technology has $\mu_n C_{ox} = 400\ \mu$A/V^2, $V_t = 0.5$ V, and $V_A' = 6$ V/μm. For a transistor with $L = 0.5\ \mu$m, find r_o and the value of W that results in $g_m = 2$ mA/V at $I_D = 0.25$ mA. Also, find r_o and the required V_{GS}.

			Voltages (V)			Dimensions (μm)				
Case	**Type**	I_D **(mA)**	$\lvert V_{GS}\rvert$	$\lvert V_t\rvert$	$\lvert V_{OV}\rvert$	W	L	W/L	$k'(W/L)$	g_m**(mA/V)**
a	N	1	3	2			1			
b	N	1		0.7	0.5	50				
c	N	10			2		1			
d	N	0.5			0.5					
e	N	0.1				10	2			
f	N		1.8	0.8		40	4			
g	P	0.5						25		
h	P		3	1					0.5	
i	P	10				4000	2			
j	P	10			4					
k	P				1	30	3			
l	P				5				0.08	

7.30 For the NMOS amplifier in Fig. P7.30, replace the transistor with its T equivalent circuit, assuming $\lambda = 0$. Derive expressions for the voltage gains v_s/v_i and v_d/v_i.

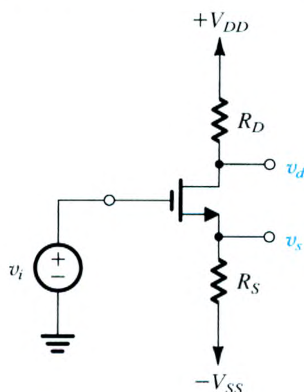


Figure P7.30

SIM **7.31** In the circuit of Fig. P7.31, the NMOS transistor has $|V_t| = 0.8$ V and $V_A = 20$ V and operates with $V_D = 1$ V. What is the voltage gain v_o/v_i? What do V_D and the gain become for I increased to 1 mA?

VE 7.1

7.32 For a 0.18-μm CMOS fabrication process: $V_{tn} = 0.5$ V, $V_{tp} = -0.5$ V, $\mu_n C_{ox} = 400\ \mu\text{A/V}^2$, $\mu_p C_{ox} = 100\ \mu\text{A/V}^2$, V_A (n-channel devices) $= 5L$ (μm), and $|V_A|$ (p-channel devices) $= 6L$ (μm). Find the small-signal model parameters g_m and r_o for both an NMOS and a PMOS transistor having $W/L =$ 5 μm/0.5 μm and operating at $I_D = 100\ \mu$A. Also, find the overdrive voltage at which each device must be operating.

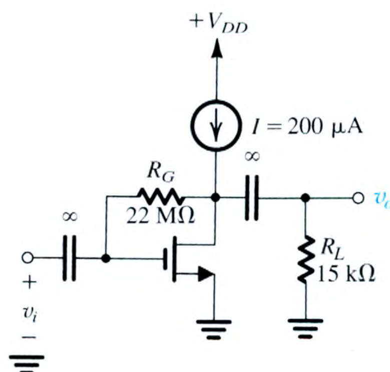


Figure P7.31

7.33 Figure P7.33 shows a discrete-circuit amplifier. The input signal v_{sig} is coupled to the gate through a very large capacitor (shown as infinite). The transistor source is connected to ground at signal frequencies via a very large capacitor (shown as infinite). The output voltage signal that develops at the drain is coupled to a load resistance via a very large capacitor (shown as infinite). All capacitors behave as short circuits for signals and as open circuits for dc.

(a) If the transistor has $V_t = 1$ V, and $k_n = 4\ \text{mA/V}^2$, verify that the bias circuit establishes $V_{GS} = 1.5$ V, $I_D = 0.5$ mA, and $V_D = +7.0$ V. That is, assume these values, and verify that they are consistent with the values of the circuit components and the device parameters.

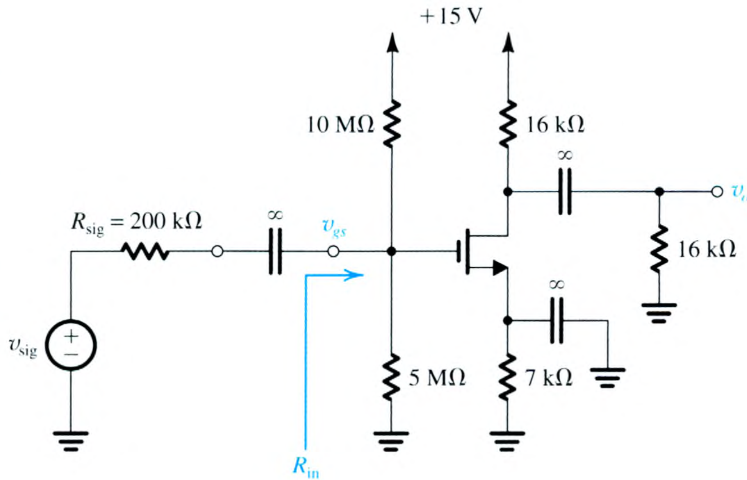


Figure P7.33

(b) Find g_m and r_o if $V_A = 100$ V.
(c) Draw a complete small-signal equivalent circuit for the amplifier, assuming all capacitors behave as short circuits at signal frequencies.
(d) Find R_{in}, v_{gs}/v_{sig}, v_o/v_{gs}, and v_o/v_{sig}.

7.34 Consider a transistor biased to operate in the active mode at a dc collector current I_C. Calculate the collector signal current as a fraction of I_C (i.e., i_c/I_C) for input signals v_{be} of +1 mV, –1 mV, +2 mV, –2 mV, +5 mV, –5 mV, +8 mV, –8 mV, +10 mV, –10 mV, +12 mV, and –12 mV. In each case do the calculation two ways:

(a) using the exponential characteristic, and
(b) using the small-signal approximation.

Present your results in the form of a table that includes a column for the error introduced by the small-signal approximation. Comment on the range of validity of the small-signal approximation.

7.35 An *npn* BJT with grounded emitter is operated with $V_{BE} = 0.700$ V, at which the collector current is 1 mA. A 2-kΩ resistor connects the collector to a +5-V supply. What is the resulting collector voltage V_C? Now, if a signal applied to the base raises v_{BE} to 705 mV, find the resulting total collector current i_C and total collector voltage v_C using the exponential i_C–v_{BE} relationship. For this situation, what are v_{be} and v_c? Calculate the voltage gain v_c/v_{be}. Compare with the value obtained using the small-signal approximation, that is, $-g_m R_C$.

7.36 A transistor with $\beta = 100$ is biased to operate at a dc collector current of 0.4 mA. Find the values of g_m, r_π, and r_e. Repeat for a bias current of 40 μA.

7.37 A *pnp* BJT is biased to operate at $I_C = 0.5$ mA. What is the associated value of g_m? If $\beta = 100$, what is the value of the small-signal resistance seen looking into the emitter (r_e)? into the base (r_π)? If the collector is connected to a 5-kΩ load, with a signal of 5-mV peak applied between base and emitter, what output signal voltage results?

D 7.38 A designer wants to create a BJT amplifier with a g_m of 20 mA/V and a base input resistance of 4000 Ω or more. What collector-bias current should she choose? What is the minimum β she can tolerate for the transistor used?

7.39 A transistor operating with nominal g_m of 40 mA/V has a β that ranges from 60 to 150. Also, the bias circuit, being less than ideal, allows a ±10% variation in I_C. What are the extreme values found of the resistance looking into the base?

7.40 In the circuit of Fig. 7.21, V_{BE} is adjusted so that $V_C = 1$ V. If $V_{CC} = 3$ V, $R_C = 2$ kΩ, and a signal $v_{be} = 0.005 \sin \omega t$ volts is applied, find expressions for the total instantaneous quantities $i_C(t)$, $v_C(t)$, and $i_B(t)$. Assume linear operation. The transistor has $\beta = 100$. What is the voltage gain?

D *7.41 We wish to design the amplifier circuit of Fig. 7.21 under the constraint that V_{CC} is fixed. Let the input signal $v_{be} = \hat{V}_{be} \sin \omega t$, where $\hat{V}_{be}$ is the maximum value for acceptable linearity. For the design that results in the largest signal at the collector, without the BJT leaving the active region, show that

$$R_C I_C = (V_{CC} - 0.3) \Big/ \left(1 + \frac{\hat{V}_{be}}{V_T}\right)$$

and find an expression for the voltage gain obtained. For $V_{CC} = 3$ V and $\hat{V}_{be} = 5$ mV, find the dc voltage at the collector, the amplitude of the output voltage signal, and the voltage gain.

7.42 The table below summarizes some of the basic attributes of a number of BJTs of different types, operating as amplifiers under various conditions. Provide the missing entries.

7.43 A BJT is biased to operate in the active mode at a dc collector current of 0.5 mA. It has a β of 100 and V_A of 100 V. Give the four small-signal models (Figs. 7.26 and 7.28) of the BJT complete with the values of their parameters.

7.44 Using the T model of Fig. 7.27(a), show that the input resistance between base and emitter, looking into the base, is equal to r_π.

7.45 Show that the collector current provided by the model of Fig. 7.27(b) is equal to that provided by the model in Fig. 7.27(a).

7.46 Show that the hybrid-π model of Fig. 7.25(b) is the incremental version of the large-signal model of Fig. 6.5(d).

7.47 Show that the T model of Fig. 7.27(b) is the incremental version of the large-signal model of Fig. 6.5(b).

7.48 The transistor amplifier in Fig. P7.48 is biased with a current source I and has a very high β. Find the dc voltage at the collector, V_C. Also, find the value of r_e. Replace the transistor with the T model of Fig. 7.27(b) (note that the dc current source I should be replaced with an open circuit). Hence find the voltage gain v_c/v_i.

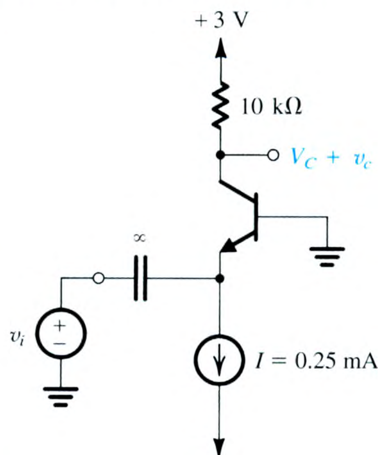


Figure P7.48

7.49 For the conceptual circuit shown in Fig. 7.24, $R_C = 5$ kΩ, $g_m = 20$ mA/V, and $\beta = 100$. If a peak-to-peak output voltage of 0.5 V is measured at the collector, what are the peak-to-peak values of v_{be} and i_b?

7.50 Figure P7.50 shows the circuit of an amplifier fed with a signal source v_{sig} with a source resistance R_{sig}. The bias

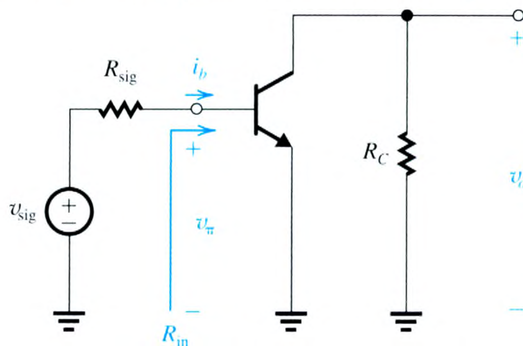


Figure P7.50

Transistor	a	b	c	d	e	f	g
α	1.000					0.90	
β		100		∞			
I_C (mA)	1.00		1.00				
I_E (mA)		1.00				5	
I_B (mA)			0.020				1.10
g_m (mA/V)							700
r_e (Ω)				25	100		
r_π (Ω)					10.1 kΩ		

circuitry is not shown. Replace the BJT with its hybrid-π equivalent circuit of Fig. 7.25(a). Find the input resistance $R_{in} \equiv v_\pi / i_b$, the voltage transmission from source to amplifier input, v_π / v_{sig}, and the voltage gain from base to collector, v_o / v_π. Use these to show that the overall voltage gain v_o / v_{sig} is given by

$$\frac{v_o}{v_{sig}} = -\frac{\beta R_C}{r_\pi + R_{sig}}$$

7.51 Figure P7.51 shows a transistor with the collector connected to the base. The bias arrangement is not shown. Since a zero v_{BC} implies operation in the active mode, the BJT can be replaced by one of the small-signal models of Figs. 7.25 and 7.27. Use the model of Fig. 7.27(b) and show that the resulting two-terminal device, known as a diode-connected transistor, has a small-signal resistance r equal to r_e.

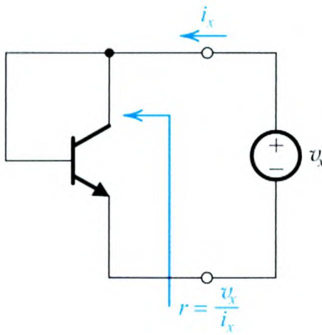


Figure P7.51

7.52 Figure P7.52 shows a particular configuration of BJT amplifiers known as "emitter follower." The bias arrangement is not shown. Replace the BJT with its T equivalent-circuit model of Fig. 7.27(b). Show that

$$R_{in} \equiv \frac{v_i}{i_b} = (\beta + 1)(r_e + R_e)$$

$$\frac{v_o}{v_i} = \frac{R_e}{R_e + r_e}$$

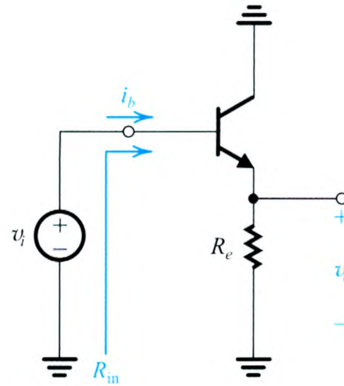


Figure P7.52

7.53 For the circuit shown in Fig. P7.53, draw a complete small-signal equivalent circuit using an appropriate T model for the BJT (use $\alpha = 0.99$). You should show the values of all components, including the model parameters. What is the input resistance R_{in}? Calculate the overall voltage gain (v_o / v_{sig}).

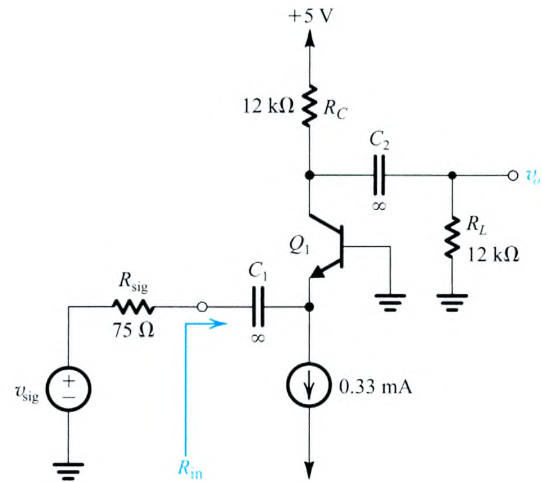


Figure P7.53

7.54 In the circuit shown in Fig. P7.54, the transistor has a β of 200. What is the dc voltage at the collector? Replacing the BJT with one of the hybrid-π models (neglecting r_o), draw the equivalent circuit of the amplifier. Find the input resistances R_{ib} and R_{in} and the overall voltage gain (v_o / v_{sig}). For an output signal of ± 0.3 V, what values of v_{sig} and v_b are required?

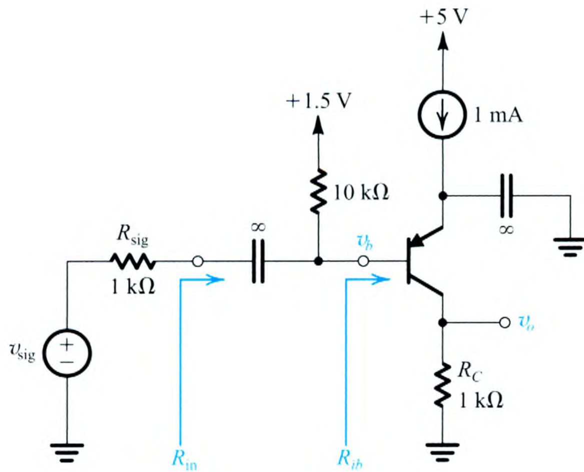


Figure P7.54

7.55 Consider the augmented hybrid-π model shown in Fig. 7.26(a). Disregarding how biasing is done, what is the largest possible voltage gain available for a signal source connected directly to the base and a very-high-resistance load? Calculate the value of the maximum possible gain for $V_A = 20$ V and $V_A = 120$ V.

D 7.56 Redesign the circuit of Fig. 7.31(a) by raising the resistor values by a factor of n to increase the resistance seen by the input v_i to 75 Ω. What value of voltage gain results? Grounded-base circuits of this kind are used in systems such as cable TV, in which, for highest-quality signaling, load resistances need to be "matched" to the equivalent resistances of the interconnecting cables.

D *7.57 Design an amplifier using the configuration of Fig. 7.31(a). The power supplies available are ±5 V. The input signal source has a resistance of 100 Ω, and the amplifier input resistance must match this value. (Note that $R_{in} = r_e \| R_E \simeq r_e$.) The amplifier is to have the greatest possible voltage gain and the largest possible output signal but retain small-signal linear operation by keeping the signal component across the base–emitter junction to no more than 5 mV. Find appropriate values for R_E and R_C. What is the value of voltage gain realized from signal source to output?

***7.58** The transistor in the circuit shown in Fig. P7.58 is biased to operate in the active mode. Assuming that β is very large, find the collector bias current I_C. Replace the transistor with the small-signal equivalent-circuit model of Fig. 7.27(b) (remember to replace the dc power supply with a short circuit). Analyze the resulting amplifier equivalent circuit to show that

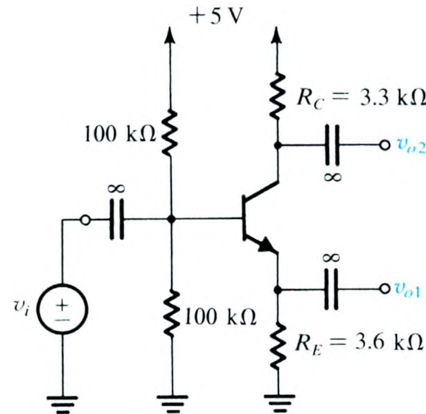


Figure P7.58

$$\frac{v_{o1}}{v_i} = \frac{R_E}{R_E + r_e}$$

$$\frac{v_{o2}}{v_i} = \frac{-\alpha R_C}{R_E + r_e}$$

Find the values of these voltage gains (for $\alpha \simeq 1$). Now, if the terminal labeled v_{o1} is connected to ground, what does the voltage gain v_{o2}/v_i become?

Section 7.3: Basic Configurations

7.59 An amplifier with an input resistance of 200 kΩ, an open-circuit voltage gain of 100 V/V, and an output resistance of 200 Ω is connected between a 20-kΩ signal source and a 2-kΩ load. Find the overall voltage gain G_v. Also find the current gain, defined as the ratio of the load current to the current drawn from the signal source.

D 7.60 Specify the parameters R_{in}, A_{vo}, and R_o of an amplifier that is to be connected between a 10-kΩ source and a 1-kΩ load and is required to meet the following specifications:

(a) No more than 5% of the signal strength is lost in the connection to the amplifier input;
(b) If the load resistance changes from the nominal value of 1 kΩ to a low value of 0.5 kΩ, the change in output voltage is limited to 5% of nominal value; and
(c) The nominal overall voltage gain is 100 V/V.

7.61 An amplifier has an input resistance of 100 kΩ, a short-circuit transconductance of 10 mA/V, and an output resistance of 100 kΩ. Find the open-circuit voltage gain A_{vo}, and the overall voltage gain G_v when the amplifier is fed with

a voltage source having $R_{sig} = 100\ \text{k}\Omega$, and a load resistance $R_L = 100\ \text{k}\Omega$ is connected at the output.

7.62 An alternative equivalent circuit of an amplifier fed with a signal source (v_{sig}, R_{sig}) and connected to a load R_L is shown in Fig. P7.62. Here G_{vo} is the open-circuit overall voltage gain,

$$G_{vo} = \left.\frac{v_o}{v_{sig}}\right|_{R_L=\infty}$$

and R_{out} is the output resistance with v_{sig} set to zero. This is different from R_o. Show that

$$G_{vo} = \frac{R_i}{R_i + R_{sig}} A_{vo}$$

where $R_i = R_{in}|_{R_L=\infty}$.

Also show that the overall voltage gain is

$$G_v = G_{vo}\frac{R_L}{R_L + R_{out}}$$

***7.63** Most practical amplifiers have internal feedback that makes them non-unilateral. In such a case, R_{in} depends on R_L. To illustrate this point we show in Fig. P7.63 the equivalent circuit of an amplifier where a feedback resistance R_f models the internal feedback mechanism that is present in this amplifier. It is R_f that makes the amplifier non-unilateral. Show that

$$R_{in} = R_1 \parallel \left[\frac{R_f + (R_2 \parallel R_L)}{1 + g_m(R_2 \parallel R_L)}\right]$$

$$A_{vo} = -g_m R_2 \frac{1 - 1/(g_m R_f)}{1 + (R_2/R_f)}$$

$$R_o = R_2 \parallel R_f$$

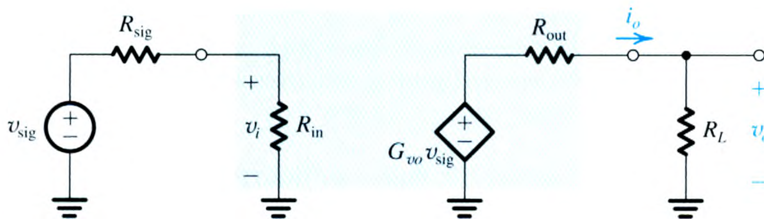


Figure P7.62

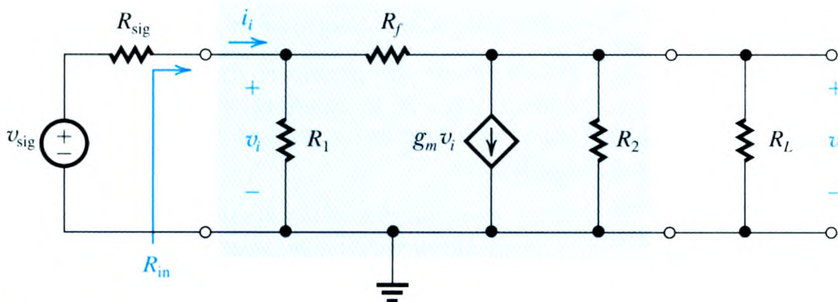


Figure P7.63

Evaluate R_{in}, A_{vo}, and R_o for the case $R_1 = 100\,\text{k}\Omega$, $R_f = 1\,\text{M}\Omega$, $g_m = 100\,\text{mA/V}$, $R_2 = 100\ \Omega$, and $R_L = 1\,\text{k}\Omega$. Which of the amplifier characteristic parameters is most affected by R_f (that is, relative to the case with $R_f = \infty$)? For $R_{sig} = 100\,\text{k}\Omega$ determine the overall voltage gain, G_v, with and without R_f present.

7.64 Calculate the overall voltage gain of a CS amplifier fed with a 1-MΩ source and connected to a 20-kΩ load. The MOSFET has $g_m = 2\,\text{mA/V}$, and a drain resistance $R_D = 20\,\text{k}\Omega$ is utilized.

7.65 A CS amplifier utilizes a MOSFET with $\mu_n C_{ox} = 400\,\mu\text{A/V}^2$ and $W/L = 10$. It is biased at $I_D = 0.5$ mA and uses $R_D = 10\,\text{k}\Omega$. Find R_{in}, A_{vo}, and R_o. Also, if a load resistance of 10 kΩ is connected to the output, what overall voltage gain G_v is realized? Now, if a 0.5-V peak sine-wave signal is required at the output, what must the peak amplitude of v_{sig} be?

D 7.66 A common-source amplifier uses a MOSFET operated at $V_{OV} = 0.2$ V. The amplifier feeds a load resistance $R_L = 10\,\text{k}\Omega$. The designer selects $R_D = 2R_L$. If you are required to realize an overall voltage gain G_v of -10 V/V what g_m do you need? Specify the bias current I_D. If, to increase the output signal swing, R_D is reduced to $R_D = R_L$, what does G_v become?

7.67 Two identical CS amplifiers are connected in cascade. The first stage is fed with a source v_{sig} having a resistance $R_{sig} = 100\,\text{k}\Omega$. A load resistance $R_L = 10\,\text{k}\Omega$ is connected to the drain of the second stage. Each MOSFET is biased at $I_D = 0.2$ mA and operates with $V_{OV} = 0.2$ V. Each stage utilizes a drain resistance $R_D = 10\,\text{k}\Omega$.

(a) Sketch the equivalent circuit of the two-stage amplifier.
(b) Calculate the overall voltage gain G_v.

7.68 A CE amplifier utilizes a BJT with $\beta = 160$ biased at $I_C = 0.4$ mA; it has a collector resistance $R_C = 10\,\text{k}\Omega$. Find R_{in}, R_o, and A_{vo}. If the amplifier is fed with a signal source with a resistance of 10 kΩ, and a load resistance $R_L = 10\,\text{k}\Omega$ is connected to the output terminal, find the resulting A_v and G_v. If the peak voltage of the sine wave appearing between base and emitter is to be limited to 5 mV, what $\hat{v}_{sig}$ is allowed, and what output voltage signal appears across the load?

D *7.69 In this problem we investigate the effect of the inevitable variability of β on the realized gain of the CE amplifier. For this purpose, use the overall gain expression in Eq. (7.114).

$$|G_v| = \frac{R_L'}{(R_{sig}/\beta) + (1/g_m)}$$

where $R_L' = R_L \,\|\, R_C$.

Consider the case $R_L' = 10\,\text{k}\Omega$ and $R_{sig} = 10\,\text{k}\Omega$, and let the BJT be biased at $I_C = 1$ mA. The BJT has a nominal β of 100.

(a) What is the nominal value of $|G_v|$?
(b) If β can be anywhere between 50 and 150, what is the corresponding range of $|G_v|$?
(c) If in a particular design, it is required to maintain $|G_v|$ within ±20% of its nominal value, what is the maximum allowable range of β?
(d) If it is not possible to restrict β to the range found in (c), and the designer has to contend with β in the range 50 to 150, what value of bias current I_C would result in $|G_v|$ falling in a range of ±20% of a new nominal value? What is the nominal value of $|G_v|$ in this case?

7.70 Two identical CE amplifiers are connected in cascade. The first stage is fed with a source v_{sig} having a resistance $R_{sig} = 10\,\text{k}\Omega$. A load resistance $R_L = 10\,\text{k}\Omega$ is connected to the collector of the second stage. Each BJT is biased at $I_C = 0.2$ mA and has $\beta = 120$. Each stage utilizes a collector resistance $R_C = 10\,\text{k}\Omega$.

(a) Sketch the equivalent circuit of the two-stage amplifier.
(b) Find the overall voltage gain, v_{o2}/v_{sig}.

7.71 A MOSFET connected in the CS configuration has a transconductance $g_m = 5\,\text{mA/V}$. When a resistance R_s is connected in the source lead, the effective transconductance is reduced to 2.5 mA/V. Estimate the value of R_s.

D 7.72 A CS amplifier using an NMOS transistor with $g_m = 4\,\text{mA/V}$ is found to have an overall voltage gain of -20 V/V. What value should a resistance R_s inserted in the source lead have to reduce the overall voltage gain to -10 V/V?

7.73 The overall voltage gain of a CS amplifier with a resistance $R_s = 200\ \Omega$ in the source lead is measured and found to be -10 V/V. When R_s is shorted, but the circuit operation remains linear, the gain doubles. What must g_m and R_D be? What value of R_s is needed to obtain an overall voltage gain of -16 V/V?

7.74 A CE amplifier uses a BJT with $\beta = 100$ biased at $I_C = 0.5$ mA and has a collector resistance $R_C = 15\,\text{k}\Omega$ and a resistance $R_e = 200\,\Omega$ connected in the emitter. Find R_{in}, A_{vo}, and R_o. If the amplifier is fed with a signal source having

a resistance of 10 kΩ, and a load resistance $R_L = 15$ kΩ is connected to the output terminal, find the resulting A_v and G_v. If the peak voltage of the sine wave appearing between base and emitter is to be limited to 5 mV, what $\hat{v}_{sig}$ is allowed, and what output voltage signal appears across the load?

D 7.75 Design a CE amplifier with a resistance R_e in the emitter to meet the following specifications:

(i) Input resistance $R_{in} = 50$ kΩ.
(ii) When fed from a signal source with a peak amplitude of 0.1 V and a source resistance of 50 kΩ, the peak amplitude of v_π is 5 mV.

Specify R_e and the bias current I_C. The BJT has $\beta = 74$. If the total resistance in the collector is 10 kΩ, find the overall voltage gain G_v and the peak amplitude of the output signal v_o.

SIM **D 7.76** Including an emitter resistance R_e reduces the variability of the gain G_v due to the inevitable wide variance in the value of β. Consider a CE amplifier operating between a signal source with $R_{sig} = 10$ kΩ and a total collector resistance $R_C \| R_L$ of 10 kΩ. The BJT is biased at $I_C = 1$ mA and its β is specified to be nominally 100 but can lie in the range of 50 to 150. First determine the nominal value and the range of $|G_v|$ without resistance R_e. Then select a value for R_e that will ensure that $|G_v|$ be within ±20% of its new nominal value. Specify the value of R_e, the new nominal value of $|G_v|$, and the expected range of $|G_v|$.

7.77 A CG amplifier using an NMOS transistor for which $g_m = 5$ mA/V has a 5-kΩ drain resistance R_D and a 5-kΩ load resistance R_L. The amplifier is driven by a voltage source with a 250-Ω resistance. What is the input resistance of the amplifier? What is the overall voltage gain G_v ? By what factor must the bias current I_D of the MOSFET be changed so that R_{in} matches R_{sig}?

7.78 A CG amplifier when fed with a signal source having $R_{sig} = 100$ Ω is found to have an overall voltage gain of 12 V/V. When a 100-Ω resistance is added in series with the signal generator, the overall voltage gain decreases to 10 V/V. What must g_m of the MOSFET be? If the MOSFET is biased at $I_D = 0.25$ mA, at what overdrive voltage must it be operating?

D 7.79 A CB amplifier is operating with $R_L = 5$ kΩ, $R_C = 5$ kΩ, and $R_{sig} = 50$ Ω. At what current I_C should the transistor be biased for the input resistance R_{in} to equal that of the signal source? What is the resulting overall voltage gain? Assume $\alpha \simeq 1$.

7.80 For the circuit in Fig. P7.80, let $R_{sig} \gg r_e$ and $\alpha \simeq 1$. Find v_o.

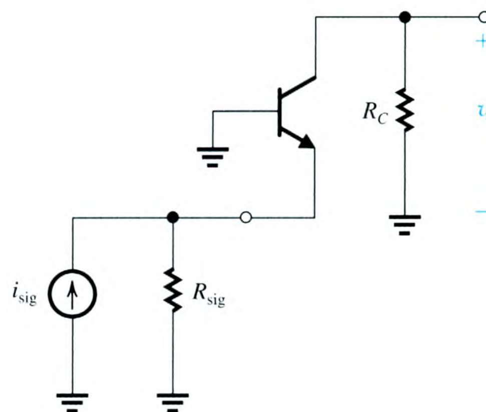


Figure P7.80

7.81 A CB amplifier biased at $I_E = 0.2$ mA with $R_C = R_L = 10$ kΩ is driven by a signal source with $R_{sig} = 0.5$ kΩ. Find the overall voltage gain G_v. If the maximum signal amplitude of the voltage between base and emitter is limited to 10 mV, what are the corresponding amplitudes of v_{sig} and v_o? Assume $\alpha \simeq 1$.

7.82 A source follower is required to connect a high-resistance source to a load whose resistance is nominally 2 kΩ but can be as low as 1.5 kΩ and as high as 5 kΩ. What is the maximum output resistance that the source follower must have if the output voltage is to remain within ±10% of nominal value? If the MOSFET has $k_n = 10$ mA/V^2, at what current I_D must it be biased? At what overdrive voltage is the MOSFET operating?

D 7.83 A source follower is required to deliver a 0.5-V peak sinusoid to a 2-kΩ load. If the peak amplitude of v_{gs} is to be limited to 50 mV, and the MOSFET transconductance parameter k_n is 5 mA/V^2, what is the lowest value of I_D at which the MOSFET can be biased? At this bias current, what are the maximum and minimum currents that the MOSFET will be conducting (at the positive and negative peaks of the output sine wave)? What must the peak amplitude of v_{sig} be?

problems with blue numbers are considered essential; * = difficult problem; ** = more difficult; *** = very challenging
SIM = simulation; D = design problem; ● = see related video example

D 7.84 An emitter follower is required to deliver a 0.5-V peak sinusoid to a 2-kΩ load. If the peak amplitude of v_{be} is to be limited to 5 mV, what is the lowest value of I_E at which the BJT can be biased? At this bias current, what are the maximum and minimum currents that the BJT will be conducting (at the positive and negative peaks of the output sine wave)? If the resistance of the signal source is 200 kΩ, what value of G_v is obtained? Thus determine the required amplitude of v_{sig}. Assume $\beta = 100$.

7.85 An emitter follower with a BJT biased at $I_C = 5$ mA and having $\beta = 200$ is connected between a source with $R_{sig} =$ 10 kΩ and a load $R_L = 200$ Ω.

(a) Find R_{in}, v_b/v_{sig}, and v_o/v_{sig}.
(b) If the signal amplitude across the base–emitter junction is to be limited to 10 mV, what is the corresponding amplitude of v_{sig} and v_o?
(c) Find the open-circuit voltage gain G_{vo} and the output resistance R_{out}. Use these values first to verify the value of G_v obtained in (a), then to find the value of G_v obtained with R_L reduced to 150 Ω.

7.86 An emitter follower, when driven from a 10-kΩ source, was found to have an output resistance R_{out} of 125 Ω. The output resistance increased to 225 Ω when the source resistance was doubled. Find the overall voltage gain when the follower is driven by a 10-kΩ source and loaded with a 1-kΩ resistor.

7.87 For the Darlington follower in Fig. 7.48(b) let Q_2 be biased at a collector current of 10 mA and let $\beta_1 = \beta_2 = 100$. If $R_{sig} = 1$ MΩ and $R_L = 1$ kΩ, find G_{vo}, R_{out}, and G_v.

7.88 For the general amplifier circuit shown in Fig. P7.88 neglect the Early effect.

(a) Find expressions for v_c/v_{sig} and v_e/v_{sig}.
(b) If v_{sig} is disconnected from node X, node X is grounded, and node Y is disconnected from ground and connected to v_{sig}, find the new expression for v_c/v_{sig}.

7.89 When the Early effect is neglected, the overall voltage gain of a CE amplifier with a collector resistance $R_C =$ 10 kΩ is calculated to be −100 V/V. If the BJT is biased at $I_C = 1$ mA and the Early voltage is 125 V, provide a better estimate of the voltage gain G_v.

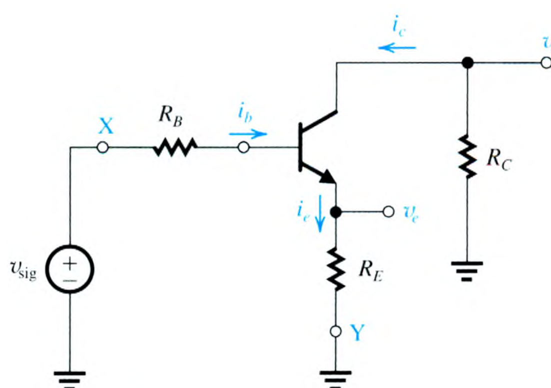


Figure P7.88

7.90 Show that when r_o is taken into account, the voltage gain of the source follower becomes

$$G_v \equiv \frac{v_o}{v_{sig}} = \frac{R_L \| r_o}{(R_L \| r_o) + \dfrac{1}{g_m}}$$

Now, with R_L removed, the voltage gain is carefully measured and found to be 0.98. Then, when R_L is connected and its value is varied, it is found that the gain is halved at $R_L = 500$ Ω. If the amplifier remained linear throughout this measurement, what must the values of g_m and r_o be?

D *7.91 In this problem, we investigate the effect of changing the bias current I_C on the overall voltage gain G_v of a CE amplifier. Consider a CE amplifier operating with a signal source $R_{sig} = 10$ kΩ and with $R_C \| R_L = 10$ kΩ. The BJT is specified to have $\beta = 100$ and $V_A = 25$ V. Use Eq. (7.114) (with r_o included in parallel with R_C and R_L in the numerator) to find $|G_v|$ at $I_C = 0.1$ mA, 0.2 mA, 0.5 mA, 1.0 mA, and 1.25 mA. Observe the effect of r_o on limiting $|G_v|$ as I_C is increased. Find the value of I_C that results in $|G_v| = 50$ V/V.

Section 7.4: Biasing



VE 7.2

D 7.92 Consider the classical biasing scheme shown in Fig. 7.50(c), using a 9-V supply. For the MOSFET, $V_t = 0.7$ V, $\lambda = 0$, and $k_n = 5$ mA/V^2. Arrange that the drain current is 0.2 mA, with about one-third of the supply voltage across each of R_S and R_D. Use 22 MΩ for the larger of R_{G1} and R_{G2}. What are the values of R_{G1}, R_{G2}, R_S, and R_D that you have chosen? Specify them to two significant digits. For your design, how far is the drain voltage from the edge of saturation?

D 7.93 Using the circuit topology displayed in Fig. 7.50(e), arrange to bias the NMOS transistor at $I_D = 0.2$ mA with V_D midway between cutoff and the beginning of triode operation. The available supplies are ±5 V. For the NMOS transistor, $V_t = 0.7$ V, $\lambda = 0$, and $k_n = 5$ mA/V^2. Use a gate-bias resistor of 10 MΩ. Specify R_S and R_D to two significant digits.

D *7.94 In an electronic instrument using the biasing scheme shown in Fig. 7.50(c), a manufacturing error reduces R_S to zero. Let $V_{DD} = 15$ V, $R_{G1} = 10$ MΩ, and $R_{G2} = 5.1$ MΩ. What is the value of V_G created? If supplier specifications allow k_n to vary from 0.2 to 0.3 mA/V^2 and V_t to vary from 1.0 V to 1.5 V, what are the extreme values of I_D that may result? What value of R_S should have been installed to limit the maximum value of I_D to 1.5 mA? Choose an appropriate standard 5% resistor value (refer to Appendix J). What extreme values of current now result?

7.95 An NMOS transistor is connected in the bias circuit of Fig. 7.50(c), with $V_G = 4$ V and $R_S = 15$ kΩ. The transistor has $V_t = 0.7$ V and $k_n = 5$ mA/V^2. What bias current results? If a transistor for which k_n is 50% higher is used, what is the resulting percentage increase in I_D?

SIM 7.96 The bias circuit of Fig. 7.50(c) is used in a design with $V_G = 5$ V and $R_S = 2$ kΩ. For a MOSFET with $k_n = 2$ mA/V^2, the source voltage is 2 V. What must V_t be for this device? In a device for which V_t is 0.5 V less, what does V_S become? What bias current results?

D 7.97 Design the circuit of Fig. 7.50(e) for a MOSFET having $V_t = 0.6$ V and $k_n = 5$ mA/V^2. Let $V_{DD} = V_{SS} = 3$ V. Design for a dc bias current of 0.4 mA and for the largest possible voltage gain (and thus the largest possible R_D) consistent with allowing a 1.2-V peak-to-peak voltage swing at the drain. Assume that the signal voltage on the source terminal of the FET is zero.

SIM D 7.98 Design the circuit in Fig. P7.98 so that the transistor operates in saturation with V_D biased 1 V from the edge of the triode region, with $I_D = 1$ mA and $V_D = 3$ V, for each of the following two devices (use a 10-μA current in the voltage divider):

(a) $|V_t| = 1$ V and $k_p' W/L = 0.5$ mA/V^2
(b) $|V_t| = 2$ V and $k_p' W/L = 1.25$ mA/V^2

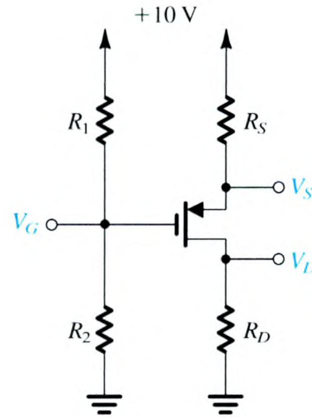


Figure P7.98

For each case, specify the values of V_G, V_D, V_S, R_1, R_2, R_S, and R_D.

D **7.99 A very useful way to characterize the stability of the bias current I_D is to evaluate the sensitivity of I_D relative to a particular transistor parameter whose variability might be large. The sensitivity of I_D relative to the MOSFET parameter $K \equiv \frac{1}{2}k'(W/L)$ is defined as

$$S_K^{I_D} \equiv \frac{\partial I_D / I_D}{\partial K / K} = \frac{\partial I_D}{\partial K}\frac{K}{I_D}$$

and its value, when multiplied by the variability (or tolerance) of K, provides the corresponding expected variability of I_D,

$$\frac{\Delta I_D}{I_D} = S_K^{I_D}\left(\frac{\Delta K}{K}\right)$$

The purpose of this problem is to investigate the use of the sensitivity function in the design of the bias circuit of Fig. 7.50(e).

(a) Show that when V_t is constant,

$$S_K^{I_D} = 1\Big/\left(1 + 2\sqrt{KI_D}R_S\right)$$

(b) For a MOSFET having $K = 100$ μA/V^2 with a variability of ±10% and $V_t = 1$ V, find the value of R_S that would result in $I_D = 100$ μA with a variability of ±1%. Also, find V_{GS} and the required value of V_{SS}.

(c) Using the same MOSFET as in (a), if the available supply $V_{SS} = 5$ V, find the value of R_S for $I_D = 100$ μA. Evaluate the sensitivity function, and give the expected variability of I_D in this case.

D *7.100 The variability ($\Delta I_D/I_D$) in the bias current I_D due to the variability ($\Delta V_t/V_t$) in the threshold voltage V_t can be evaluated from

$$\frac{\Delta I_D}{I_D} = S_{V_t}^{I_D}\left(\frac{\Delta V_t}{V_t}\right)$$

where $S_{V_t}^{I_D}$, the sensitivity of I_D relative to V_t, is defined as

$$S_{V_t}^{I_D} = \frac{\partial I_D}{\partial V_t}\frac{V_t}{I_D}$$

(a) For a MOSFET biased with a fixed V_{GS}, show that

$$S_{V_t}^{I_D} = -\frac{2V_t}{V_{OV}}$$

and find the variability in I_D for $V_t = 0.5$ V and $\Delta V_t/V_t = \pm 5\%$. Let the MOSFET be biased at $V_{OV} = 0.25$ V.

(b) For a MOSFET biased with a fixed gate voltage V_G and a resistance R_S included in the source lead, show that

$$S_{V_t}^{I_D} = -\frac{2V_t}{V_{OV} + 2I_D R_S}$$

For the same parameters given in (a), find the required value of ($I_D R_S$) and V_G to limit $\Delta I_D/I_D$ to $\pm 5\%$. What value of R_S is needed if I_D is 100 μA?

SIM **7.101** In the circuit of Fig. 7.52, let R_G = 10 MΩ, R_D = 5 kΩ, and V_{DD} = 5 V. For each of the following two transistors, find the voltages V_D and V_G.

(a) $V_t = 0.7$ V and $k_n = 5$ mA/V^2
(b) $V_t = 1.5$ V and $k_n = 10$ mA/V^2

D 7.102 Using the feedback bias arrangement shown in Fig. 7.52 with a 3-V supply and an NMOS device for which $V_t = 0.75$ V and $k_n = 3.2$ mA/V^2, find R_D to establish a drain current of 0.2 mA.

D 7.103 Figure P7.103 shows a variation of the feedback-bias circuit of Fig. 7.52. Using a 3-V supply with an NMOS transistor for which V_t = 0.8 V, k_n = 4 mA/V^2, and λ = 0, provide a design that biases the transistor at I_D = 0.5 mA, with V_{DS} large enough to allow saturation operation for a 1-V negative signal swing at the drain. Use 13 MΩ as the largest resistor in the feedback-bias network. What values of R_D, R_{G1}, and R_{G2} have you chosen?

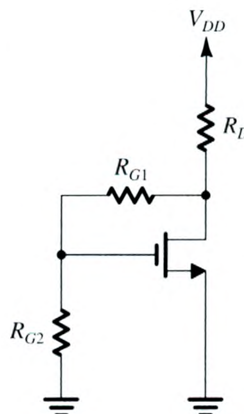


Figure P7.103

D 7.104 For the circuit in Fig. 7.53(a), neglect the base current I_B in comparison with the current in the voltage divider. It is required to bias the transistor at I_C = 1 mA, which requires selecting R_{B1} and R_{B2} so that V_{BE} = 0.710 V. If V_{CC} = 3 V, what must the ratio R_{B1}/R_{B2} be? Now, if R_{B1} and R_{B2} are 1% resistors, that is, each can be in the range of 0.99 to 1.01 of its nominal value, what is the range obtained for V_{BE}? What is the corresponding range of I_C? If R_C = 2 kΩ, what is the range obtained for V_{CE}? Comment on the efficacy of this biasing arrangement.

D 7.105 It is required to bias the transistor in the circuit of Fig. 7.53(b) at I_C = 1 mA. The transistor β is specified to be nominally 100, but it can fall in the range of 50 to 150. For V_{CC} = +3 V and R_C = 2 kΩ, find the required value of R_B to achieve I_C = 1 mA for the "nominal" transistor. What is the expected range for I_C and V_{CE}? Comment on the efficacy of this bias design.

D 7.106 Consider the single-supply bias network in Fig. 7.54(a). Provide a design using a 9-V supply in which the supply voltage is equally split between R_C, V_{CE}, and R_E with a collector current of 0.5 mA. The transistor β is specified to have a minimum value of 80. Use a voltage-divider current of $I_E/10$, or slightly higher. Since a reasonable design should operate for the best transistors for which β is very high, do your initial design with $\beta = \infty$. Then choose the 5% resistors (see Appendix J) required to produce a V_{BB} that is slightly higher than the ideal value. Specify the values you have chosen for R_E, R_C, R_1, and R_2. Now, find V_B, V_E, V_C, and I_C for your final design using $\beta = 80$.

D 7.107 This problem investigates an alternative solution to the design problem specified in Problem 7.106. You are required to design the bias circuit in Fig. 7.54(a) for the case $V_{CC}=9$ V. Design for the supply voltage equally split between R_C, V_{CE}, and R_E with a collector current of $I_C=0.5$ mA. Your design should be based on a minimum value of 80 for the transistor β. Use a current in R_1 equal to $I_C/10$. Begin your design by finding $R_E=V_E/I_E$ and then find R_C, R_1, and R_2. Choose suitable 5% resistors (see Appendix J). Specify the values you have chosen for R_E, R_C, R_1, and R_2, and find I_E, I_C, V_B, and V_C.

D *7.108 You are required to design the bias circuit of Fig. 7.54(a) for a BJT whose nominal $\beta=100$.

(a) Find the largest ratio (R_B/R_E) that will guarantee I_E remains within ±5% of its nominal value for β as low as 50 and as high as 150.
(b) If the resistance ratio found in (a) is used, find an expression for the voltage $V_{BB}\equiv V_{CC}R_2/(R_1+R_2)$ that will result in a voltage drop of $V_{CC}/3$ across R_E.
(c) For $V_{CC}=5$ V, find the required values of R_1, R_2, and R_E to obtain $I_E=0.5$ mA and to satisfy the requirement for stability of I_E in (a).
(d) Find R_C so that $V_{CE}=1.0$ V for β equal to its nominal value.

Check your design by evaluating the resulting range of I_E.

D 7.109 Design the circuit in Fig. 7.55 using ±1.5-V supplies, $I_C=0.2$ mA, and V_C placed 1 V above V_E. The transistor to be used has a minimum β of 50. Design for the device with the lowest β and use the largest value for R_B consistent with limiting the voltage drop across it to approximately one-tenth the voltage drop across R_E. Use standard 5% resistors (see Appendix J). Evaluate your design by calculating I_C, V_E, V_B, and V_C for the two cases: $\beta=\infty$ and $\beta=50$.

D *7.110 Using ±3-V power supplies, design a version of the circuit in Fig. 7.55 in which the signal will be coupled to the emitter and thus R_B can be set to zero. Find values for R_E and R_C so that a dc emitter current of 0.2 mA is obtained and so that the gain is maximized while allowing ±1 V of signal swing at the collector. Assume $\beta=50$. If temperature increases from the nominal value of 25°C to 125°C, estimate the percentage change in collector bias current. In addition to the –2 mV/°C change in V_{BE}, assume that the transistor β changes over this temperature range from 50 to 150.

SIM **D 7.111** Using a 3-V power supply, design a version of the circuit of Fig. 7.56 to provide a dc emitter current of 0.5 mA and to allow a ±1-V signal swing at the collector. The BJT has a nominal $\beta=100$. Use standard 5% resistor values (see Appendix J). If the actual BJT used has $\beta=50$, what emitter current is obtained? Also, what is the allowable signal swing at the collector? Repeat for $\beta=150$.

D *7.112 (a) Using a 3-V power supply, design the feedback bias circuit of Fig. 7.56 to provide $I_C=1$ mA and $V_C=V_{CC}/2$ for $\beta=100$.
(b) Select standard 5% resistor values, and reevaluate V_C and I_C for $\beta=100$.
(c) Find V_C and I_C for $\beta=\infty$.
(d) To improve the situation that occurs when high-β transistors are used, we have to arrange for an additional current to flow through R_B. We can do this by connecting a resistor between base and emitter, as shown in Fig. P7.112.

Design this circuit for $\beta=100$. Use a current through R_{B2} equal to the base current. Now, what values of V_C and I_C result with $\beta=\infty$?

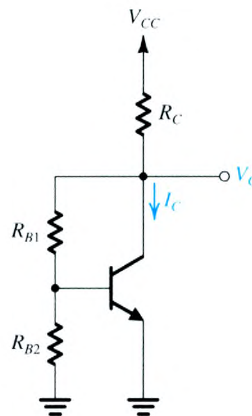


Figure P7.112

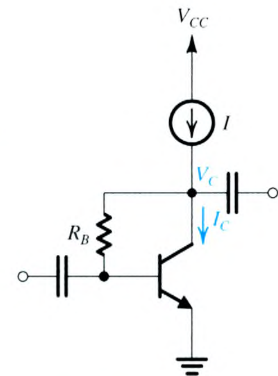


Figure P7.113

D 7.113 A circuit that can provide a very large voltage gain for a high-resistance load is shown in Fig. P7.113. Find the values of I and R_B to bias the BJT at $I_C=0.5$ mA and $V_C=1.5$ V. Let $\beta=100$.

7.114 The circuit in Fig. P7.114 provides a constant current I_O as long as the circuit to which the collector is connected maintains the BJT in the active mode. Show that

$$I_O = \alpha \frac{V_{CC}[R_2/(R_1 + R_2)] - V_{BE}}{R_E + (R_1 \parallel R_2)/(\beta + 1)}$$

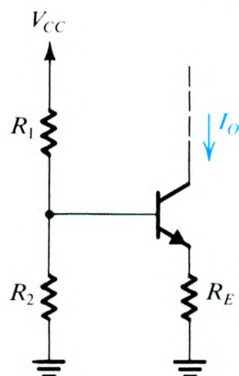


Figure P7.114

D *7.115 For the circuit in Fig. P7.115, assuming all transistors to be identical with β infinite, derive an expression for the output current I_O, and show that by selecting $R_1 = R_2$ and keeping the currents in all junctions equal, the current I_O will be

$$I_O = \frac{V_{CC}}{2R_E}$$

which is independent of V_{BE}. What must the relationship of R_E to R_1 and R_2 be? For $V_{CC} = 10$ V and $V_{BE} = 0.7$ V, design the circuit to obtain an output current of 0.5 mA. What is the lowest voltage that can be applied to the collector of Q_3?

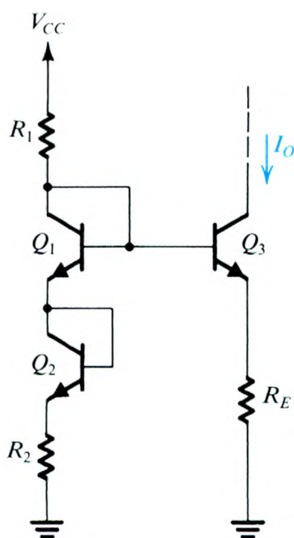


Figure P7.115

D 7.116 For the circuit in Fig. P7.116 find the value of R that will result in $I_O \simeq 0.5$ mA. What is the largest voltage that can be applied to the collector? Assume $|V_{BE}| = 0.7$ V.

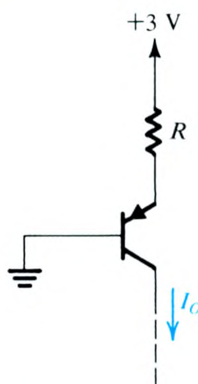


Figure P7.116

Section 7.5: Discrete-Circuit Amplifiers

7.117 Calculate the overall voltage gain G_v of a common-source amplifier for which $g_m = 5$ mA/V, $r_o = 50$ kΩ, $R_D = 10$ kΩ, and $R_{G1} || R_{G2} = 10$ MΩ. The amplifier is fed from a signal source with a Thévenin resistance of 1 MΩ, and the amplifier output is coupled to a load resistance of 20 kΩ.

SIM **7.118** The NMOS transistor in the CS amplifier shown in Fig. 7.57(a) has $V_t = 0.75$ V and $V_A = 50$ V. Also, $V_{DD} = 5$ V, $R_{G1} = 750$ kΩ, $R_{G2} = 500$ kΩ, $R_D = 10$ kΩ, $R_S = 4$ kΩ, $R_{sig} = 100$ kΩ, and $R_L = 10$ kΩ.

(a) Neglecting the Early effect, verify that the MOSFET is operating in saturation with $I_D = 0.25$ mA and $V_{OV} = 0.25$ V. What must the MOSFET's k_n be? What is the dc voltage at the drain?
(b) Find R_{in} and G_v.
(c) If v_{sig} is a sinusoid with a peak amplitude $\hat{v}_{sig}$, find the maximum allowable value of $\hat{v}_{sig}$ for which the transistor remains in saturation. What is the corresponding amplitude of the output voltage?
(d) What is the value of resistance R that needs to be inserted in series with capacitor C_S in order to allow us to double the input signal $\hat{v}_{sig}$? What output voltage now results?

SIM **D *7.119** The PMOS transistor in the CS amplifier of Fig. P7.119 has $V_{tp} = -0.75$ V and a very large $|V_A|$.

(a) Select a value for R_S to bias the transistor at $I_D = 0.5$ mA and $|V_{OV}| = 0.25$ V. Assume v_{sig} to have a zero dc component.
(b) Select a value for R_D that results in $G_v = -12$ V/V.

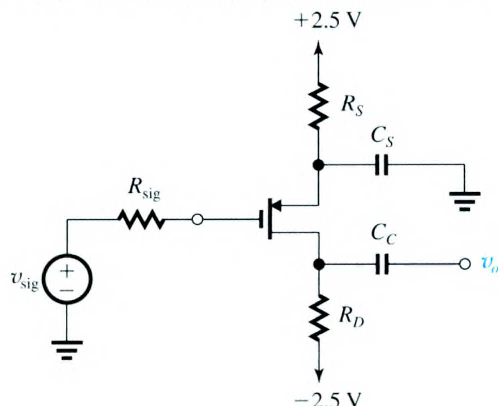


Figure P7.119

(c) Find the largest sinusoid $\hat{v}_{sig}$ that the amplifier can handle while remaining in the saturation region. What is the corresponding signal at the output?
(d) If, to obtain reasonably linear operation, $\hat{v}_{sig}$ is limited to 50 mV, what value can R_D be increased to while maintaining saturation-region operation? What is the new value of G_v?

7.120 Figure P7.120 shows a scheme for coupling and amplifying a high-frequency pulse signal. The circuit uses two MOSFETs whose bias details are not shown and a 50-Ω coaxial cable. Transistor Q_1 operates as a CS amplifier and Q_2 as a CG amplifier. For proper operation, transistor Q_2 is required to present a 50-Ω resistance to the cable. This situation is known as "proper termination" of the cable and ensures that there will be no signal reflection coming back on the cable. When the cable is properly terminated, its input resistance is 50 Ω. What must g_{m2} be? If Q_1 is biased at the same point as Q_2, what is the amplitude of the current pulses in the drain of Q_1? What is the amplitude of the voltage pulses at the drain of Q_1? What value of R_D is required to provide 1-V pulses at the drain of Q_2?

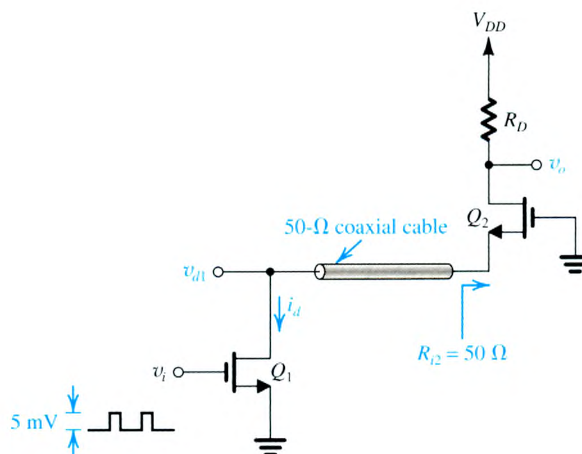


Figure P7.120

D ***7.121** The MOSFET in the circuit of Fig. P7.121 has $V_t = 0.8$ V, $k_n = 5$ mA/V^2, and $V_A = 40$ V.

(a) Find the values of R_S, R_D, and R_G so that $I_D = 0.4$ mA, the largest possible value for R_D is used while a maximum signal swing at the drain of ±0.8 V is possible, and the input resistance at the gate is 10 MΩ. Neglect the Early effect.
(b) Find the values of g_m and r_o at the bias point.
(c) If terminal Z is grounded, terminal X is connected to a signal source with a resistance of 1 MΩ, and terminal Y is connected to a load resistance of 10 kΩ, find the voltage gain from signal source to load.
(d) If terminal Y is grounded, find the voltage gain from X to Z with Z open-circuited. What is the output resistance of the source follower?
(e) If terminal X is grounded and terminal Z is connected to a current source delivering a signal current of 50 μA and having a resistance of 100 kΩ, find the voltage signal that can be measured at Y. For simplicity, neglect the effect of r_o.

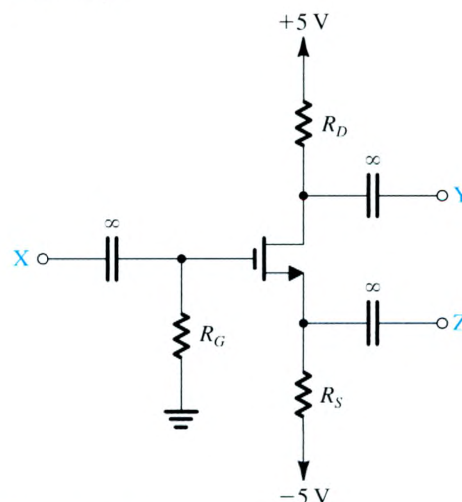


Figure P7.121

***7.122** (a) The NMOS transistor in the source-follower circuit of Fig. P7.122(a) has $g_m = 5$ mA/V and a large r_o. Find the open-circuit voltage gain and the output resistance.

(b) The NMOS transistor in the common-gate amplifier of Fig. P7.122(b) has $g_m = 5$ mA/V and a large r_o. Find the input resistance and the voltage gain.

(c) If the output of the source follower in (a) is connected to the input of the common-gate amplifier in (b), use the results of (a) and (b) to obtain the overall voltage gain v_o/v_i.

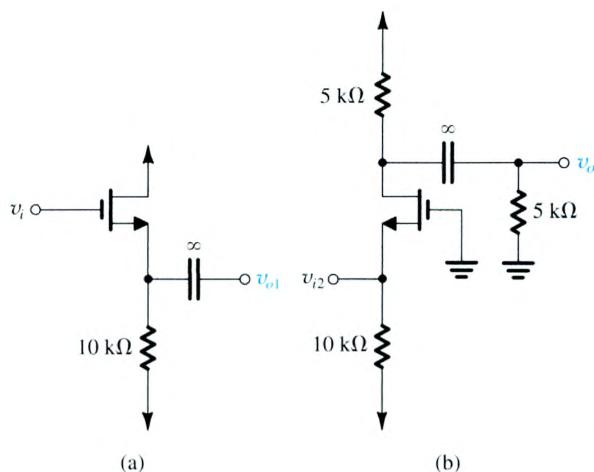


Figure P7.122

D *7.123 The MOSFET in the amplifier circuit of Fig. P7.123 has $V_t = 0.6$ V, $k_n = 5$ mA/V^2, and $V_A = 60$ V. The signal v_{sig} has a zero average.

(a) It is required to bias the transistor to operate at an overdrive voltage $V_{OV} = 0.2$ V. What must the dc voltage at the drain be? Calculate the dc drain current I_D taking into account V_A. Now, what value must the drain resistance R_D have?

(b) Calculate the values of g_m and r_o at the bias point established in (a).

(c) Using the small-signal equivalent circuit of the amplifier, show that the voltage gain is given by

$$\frac{v_o}{v_{sig}} = -\frac{R_2/R_1}{1 + \dfrac{1 + R_2/R_1}{g_m(R_D \| r_o \| R_2)(1 - 1/g_m R_2)}}$$

and find the value of the gain.

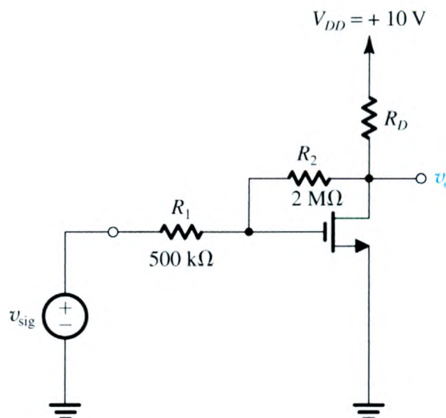


Figure P7.123

P.S. This feedback amplifier and the gain expression should remind you of an op amp utilized in the inverting configuration. We will study feedback formally in Chapter 11.

D *7.124 The MOSFET in the amplifier circuit of Fig. P7.124 has $V_t = 0.6$ V and $k_n = 5$ mA/V^2. We will assume that V_A is sufficiently large so that we can ignore the Early effect. The input signal v_{sig} has a zero average.

(a) It is required to bias the transistor to operate at an overdrive voltage $V_{OV} = 0.2$ V. What must the dc voltage at the drain be? Calculate the dc drain current I_D. What value must R_D have?

(b) Calculate the value of g_m at the bias point.

(c) Use the small-signal equivalent circuit of the amplifier to show that

$$\frac{v_o}{v_{sig}} = \frac{1 + (R_2/R_1)}{1 + \dfrac{(1 + R_2/R_1)}{g_m R_D'}}$$

and

$$R_{in} = \frac{1}{g_m}\left(1 + g_m R_D' \frac{R_1}{R_1 + R_2}\right)$$

where

$$R_D' = R_D \| (R_1 + R_2)$$

(d) Evaluate v_o/v_{sig} and R_{in}.

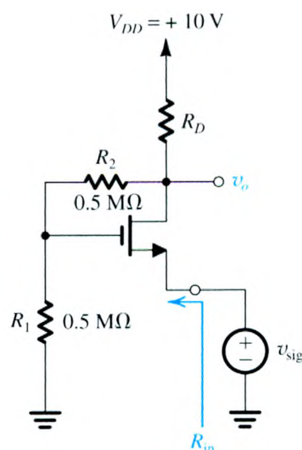


Figure P7.124

P.S. This feedback amplifier circuit and the gain formula should remind you of an op amp connected in the noninverting configuration. We will study feedback formally in Chapter 11.

7.125 For the common-emitter amplifier in Fig. 7.58(a), let $V_{CC} = 9$ V, $R_{B1} = 120$ kΩ, $R_{B2} = 62$ kΩ, $R_E = 4.7$ kΩ, and $R_C = 10$ kΩ. The transistor has $\beta = 100$. Calculate the dc bias current I_C. If the amplifier operates between a source for which $R_{sig} = 5$ kΩ and a load of 5 kΩ, replace the transistor with its hybrid-π model, and find the values of R_{in}, and the overall voltage gain v_o/v_{sig}.

D 7.126 Using the topology of Fig. 7.58(a), design an amplifier to operate between a 2-kΩ source and a 2-kΩ load with a gain v_o/v_{sig} of −40 V/V. The power supply available is 15 V. Use an emitter current of approximately 2 mA and a current of about one-tenth of that in the voltage divider that feeds the base, with the dc voltage at the base about one-third of the supply. The transistor available has $\beta = 100$. Use standard 5% resistors (see Appendix J).

D 7.127 A designer, having examined the situation described in Problem 7.125 and estimating the available gain to be approximately −30 V/V, wants to explore the possibility of improvement by reducing the loading of the source by the amplifier input. As an experiment, the designer varies the resistance levels by a factor of approximately 3: R_{B1} to 360 kΩ, R_{B2} to 180 kΩ, R_E to 15 kΩ, and R_C to 30 kΩ (standard values of 5%-tolerance resistors). With $V_{CC} = 9$ V, $R_{sig} = 5$ kΩ, $R_L = 5$ kΩ, and $\beta = 100$, what does the gain become? Comment.

D *7.128 The CE amplifier circuit of Fig. P7.128 is biased with a constant-current source I. It is required to design the circuit (i.e., find values for I, R_B, and R_C) to meet the following specifications:

(a) $R_{in} \simeq 10$ kΩ.
(b) The dc voltage drop across R_B is approximately 0.2 V.
(c) The open-circuit voltage gain from base to collector is the maximum possible, consistent with the requirement that the collector voltage never fall by more than approximately 0.4 V below the base voltage with the signal between base and emitter being as high as 5 mV.

Assume that v_{sig} is a sinusoidal source, the available supply $V_{CC} = 5$ V, and the transistor has $\beta = 100$. Use standard 5% resistance values, and specify the value of I to one significant digit. What base-to-collector open-circuit voltage gain does your design provide? If $R_{sig} = R_L = 20$ kΩ, what is the overall voltage gain?

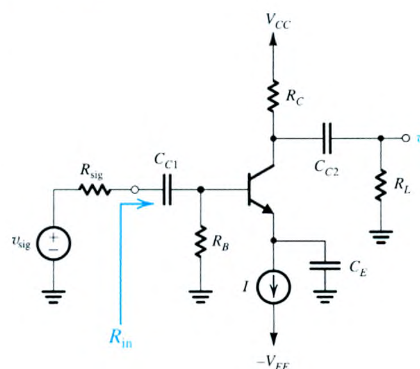


Figure P7.128

VE 7.3

D 7.129 In the circuit of Fig. P7.129, v_{sig} is a small sine-wave signal with zero average. The transistor β is 100.

(a) Find the value of R_E to establish a dc emitter current of about 0.2 mA.
(b) Find R_C to establish a dc collector voltage of about +0.5 V.
(c) For $R_L = 10$ kΩ, draw the small-signal equivalent circuit of the amplifier and determine its overall voltage gain.

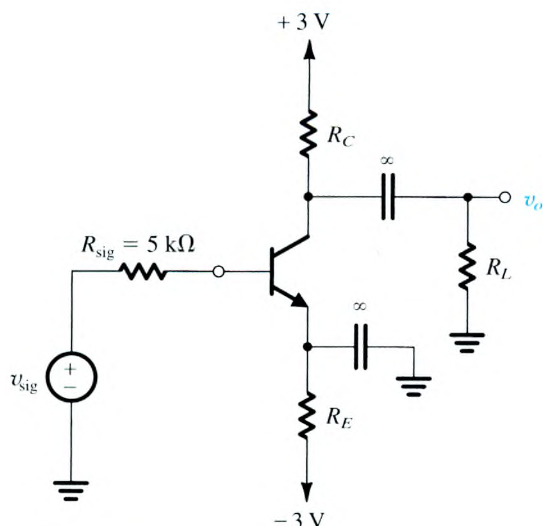




Figure P7.129

***7.130** The amplifier of Fig. P7.130 consists of two identical common-emitter amplifiers connected in cascade. Observe that the input resistance of the second stage, R_{in2}, constitutes the load resistance of the first stage.

(a) For $V_{CC} = 15$ V, $R_1 = 100$ kΩ, $R_2 = 47$ kΩ, $R_E = 3.9$ kΩ, $R_C = 6.8$ kΩ, and $\beta = 100$, determine the dc collector current and dc collector voltage of each transistor.

(b) Draw the small-signal equivalent circuit of the entire amplifier and give the values of all its components.

(c) Find R_{in1} and v_{b1}/v_{sig} for $R_{sig} = 5$ kΩ.

(d) Find R_{in2} and v_{b2}/v_{b1}.

(e) For $R_L = 2$ kΩ, find v_o/v_{b2}.

(f) Find the overall voltage gain v_o/v_{sig}.

VE 7.4

7.131 In the circuit of Fig. P7.131, the BJT is biased with a constant-current source, and v_{sig} is a small sine-wave signal. Find R_{in} and the gain v_o/v_{sig}. Assume $\beta = 100$. If the amplitude of the signal v_{be} is to be limited to 5 mV, what is the largest signal at the input? What is the corresponding signal at the output?

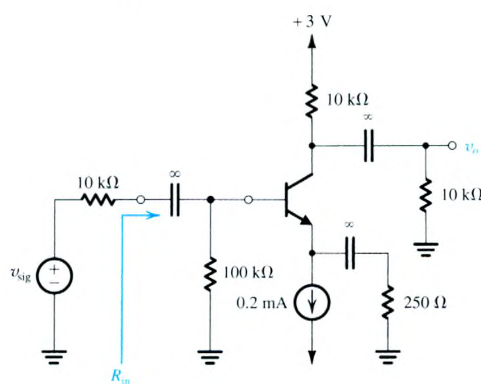


Figure P7.131

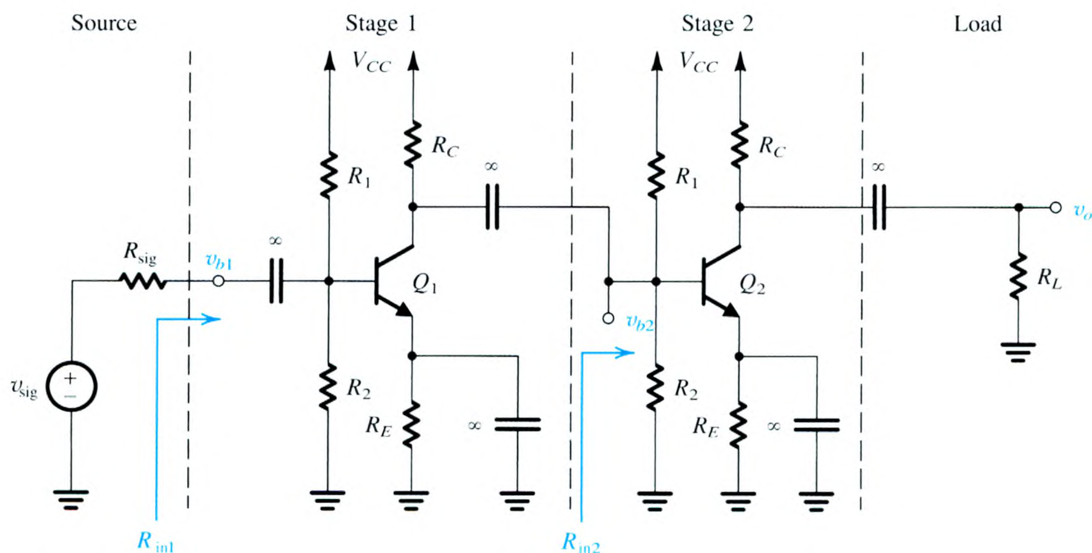


Figure P7.130

7.132 The BJT in the circuit of Fig. P7.132 has $\beta = 100$.

(a) Find the dc collector current and the dc voltage at the collector.

(b) Replacing the transistor with its T model, draw the small-signal equivalent circuit of the amplifier. Analyze the resulting circuit to determine the voltage gain v_o/v_i.

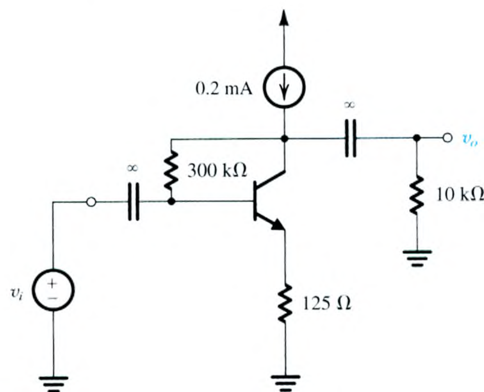


Figure P7.132

D 7.133 For the circuit in Fig. P7.133, find the value of the bias current I that results in $R_{in} = R_{sig}$. Also, find the voltage gain v_o/v_{sig}. Assume that the source provides a small signal v_{sig} and that $\beta = 100$.

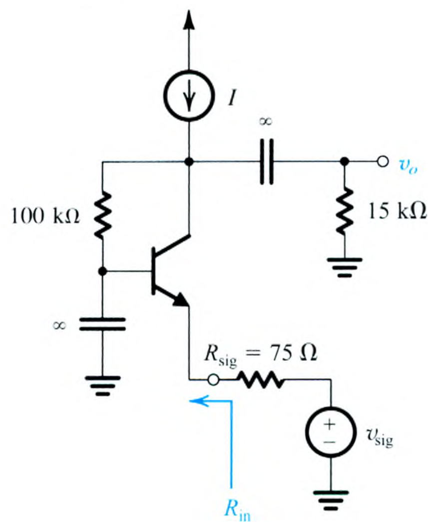


Figure P7.133

7.134 For the emitter-follower circuit shown in Fig. P7.134, the BJT used is specified to have β values in the range of 80 to 180 (a distressing situation for the circuit designer).

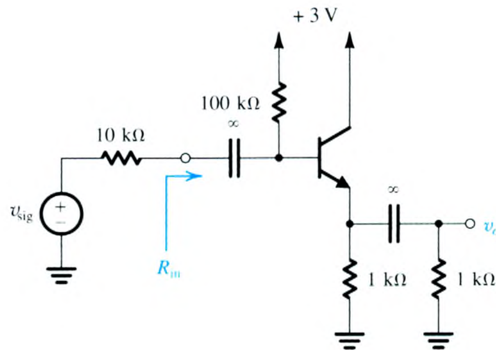


Figure P7.134

For the two extreme values of β ($\beta = 80$ and $\beta = 180$), find:

(a) I_E, V_E, and V_B

(b) the input resistance R_{in}

(c) the voltage gain v_o/v_{sig}

7.135 For the emitter follower in Fig. P7.135, the signal source is directly coupled to the transistor base. If the dc component of v_{sig} is zero, find the dc emitter current. Assume $\beta = 100$. Neglecting r_o, find R_{in}, the voltage gain v_o/v_{sig}, the current gain i_o/i_i, and the output resistance R_{out}.

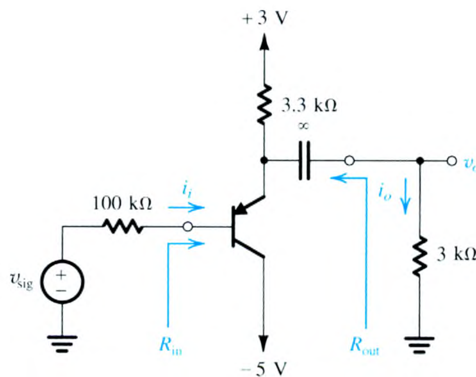


Figure P7.135

7.136 For the circuit in Fig. P7.136, called a **bootstrapped follower:

(a) Find the dc emitter current and g_m, r_e, and r_π. Use $\beta = 100$.
(b) Replace the BJT with its T model (neglecting r_o), and analyze the circuit to determine the input resistance R_{in} and the voltage gain v_o/v_{sig}.
(c) Repeat (b) for the case when capacitor C_B is open-circuited. Compare the results with those obtained in (b) to find the advantages of bootstrapping.

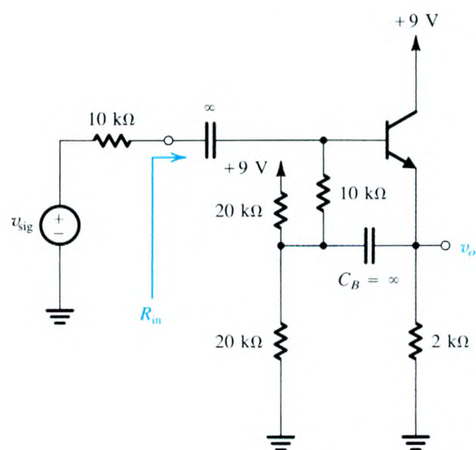


Figure P7.136

*7.137 For the follower circuit in Fig. P7.137, let transistor Q_1 have $\beta = 50$ and transistor Q_2 have $\beta = 100$, and neglect the effect of r_o. Use $V_{BE} = 0.7$ V.

(a) Find the dc emitter currents of Q_1 and Q_2. Also, find the dc voltages V_{B1} and V_{B2}.
(b) If a load resistance $R_L = 1$ kΩ is connected to the output terminal, find the voltage gain from the base to the emitter of Q_2, v_o/v_{b2}, and find the input resistance R_{ib2}

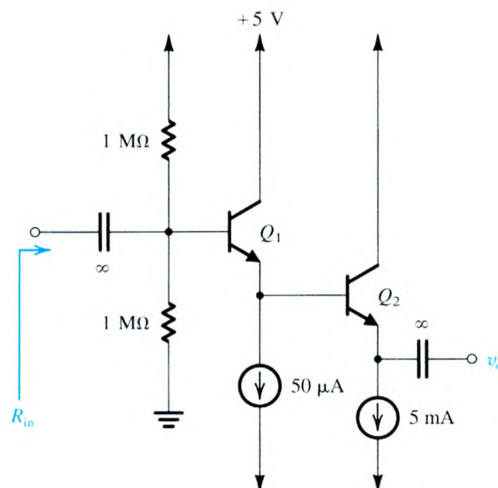


Figure P7.137

looking into the base of Q_2. (*Hint:* Consider Q_2 as an emitter follower fed by a voltage v_{b2} at its base.)
(c) Replacing Q_2 with its input resistance R_{ib2} found in (b), analyze the circuit of emitter follower Q_1 to determine its input resistance R_{in} and the gain from its base to its emitter, v_{e1}/v_{b1}.
(d) If the circuit is fed with a source having a 100-kΩ resistance, find the transmission to the base of Q_1, v_{b1}/v_{sig}.
(e) Find the overall voltage gain v_o/v_{sig}.

D 7.138 A CE amplifier has a midband voltage gain of $|A_M| = 100$ V/V, a lower 3-dB frequency $f_L = 100$ Hz, and a higher 3-dB frequency $f_H = 500$ kHz. In Chapter 10 we will learn that connecting a resistance R_e in the emitter of the BJT results in lowering f_L and raising f_H by the factor $(1 + g_m R_e)$. If the BJT is biased at $I_C = 1$ mA, find R_e that will result in f_H at least equal to 2 MHz. What will the new values of f_L and A_M be?

PART II

Analog Integrated Circuits

In Part I, we studied the MOSFET and the BJT and became familiar with their basic circuit applications. In particular, we learned how to design and analyze discrete-circuit transistor amplifiers. These amplifier circuits are assembled from discrete transistors, resistors, and capacitors on printed-circuit boards. This technology for fabricating electronic circuits and systems is economical only for low-volume production. Circuits and systems that are to be produced in high volumes (that is, millions of units) can be produced much more economically and reliably using integrated-circuit (IC) technology. Part II is devoted to the design and analysis of analog integrated circuits, while digital integrated circuits are the subject of Part III.

Integrated circuits (ICs) are expensive to produce in low volumes, and it can be costly to make the first unit due to the need for expensive fabrication masks (see Appendix A). ICs are also more difficult to debug and repair than are discrete circuits. However, once the masks are made, it is very economical to make millions of a particular IC. Also, the resulting circuits and systems are much smaller, and very complex electronic systems can be reliably made with integrated circuits. Because the lengths of the wires on an IC are so much smaller, speed and bandwidth are often increased and power is reduced.

The eight chapters of Part II constitute a coherent treatment of analog integrated circuits and can thus serve as a second course in electronic circuits. While the first six chapters (8–13) focus on the design and analysis of amplifiers, the last two chapters (14 and 15) deal with two other important analog functional blocks: filters and oscillators.

Throughout Part II, MOSFET and BJT circuits are treated side by side. Because over 90% of ICs today employ the MOSFET, its circuits are presented first. Nevertheless, BJT circuits are presented with equal depth, although sometimes somewhat more briefly. In this regard, we draw the reader's attention to Appendix G (on the website), which presents a valuable compilation of the properties of transistors of both types, allowing interesting comparisons to be made. In particular, refer to Table G.3, which is included also in the Summary Tables section on the website. As well, typical device parameter values are provided in Appendix K for a number of CMOS and bipolar fabrication technologies.

CHAPTER 8

Building Blocks of Integrated-Circuit Amplifiers

IN THIS CHAPTER YOU WILL LEARN

- The basic integrated-circuit (IC) design philosophy and how it differs from that for discrete-circuit design.
- How current sources are used to bias IC amplifiers, and how current mirrors are used to replicate the reference current generated in one location at other locations on the IC chip.
- The basic gain cells of IC amplifiers, namely, the CS and CE amplifiers with current-source loads.
- How the CG and CB amplifiers act as current buffers.
- How to increase the gain of the basic gain cell through the principle of cascoding.
- Analysis and design of the cascode amplifier and the cascode current source in both their MOS and bipolar forms.
- How the IC source follower differs from the discrete version.
- Some ingenious analog circuit-design techniques that result in current mirrors with vastly improved characteristics.

Introduction

Having studied the two major transistor types, the MOSFET and the BJT, and their basic discrete-circuit amplifier configurations, we are now ready to begin the study of integrated-circuit (IC) amplifiers. This chapter is devoted to the design of the basic building blocks of IC amplifiers.

We begin with a brief section on the design philosophy of integrated circuits and how it differs from that of discrete circuits. This is followed by the study of IC biasing in Section 8.2, highlighting the design of current sources and current mirrors. The current mirror is one of the most important building blocks of analog integrated circuits. More advanced mirror circuits are presented in Section 8.7.

The heart of this chapter is the material in Sections 8.3 to 8.5. In Section 8.3 we present the basic gain cell of IC amplifiers, namely, the current-source-loaded common-source (common-emitter) amplifier. Then, in exploring how to increase its gain, we discover the need for current buffers. In Section 8.4, we study the two amplifier configurations capable of implementing a current buffer, the common-gate and common-base amplifiers. This study differs from that in Chapter 7 in that r_o of the transistor is included, as must always be the case in integrated circuits. The study of the CG and CB leads naturally and seamlessly to the principle of cascoding and its applications in amplifier design. The cascode amplifier combines a CS (CE) gain cell with a CG (CB) current buffer to obtain higher gain and other desirable characteristics.

Throughout this chapter, MOS and bipolar circuits are presented side by side, which allows a certain economy in presentation and, more important, provides an opportunity to compare and contrast the two circuit types.

8.1 IC Design Philosophy

Integrated-circuit fabrication technology (Appendix A) imposes constraints on and provides opportunities to the circuit designer. To cope with the constraints and take advantage of the opportunities, IC designers have over the years invented (and continue to invent) many ingenious techniques, and a distinct philosophy has emerged for the design of integrated circuits. In the following we provide a brief summary of the important constraints and opportunities and the major features of the IC design philosophy:

1. In IC design, chip area is at a premium, and good designs are those that are economical in their use of "silicon real estate." This area constraint dictates the use of transistors in preference to resistors. Transistors on an IC can be made small and cheap. Thus, while the designer of discrete circuits aims to minimize the number of transistors, this is not the case in IC design.
2. Unlike the designer of discrete circuits, who is limited to available off-the-shelf transistors, the IC designer has the freedom to specify the transistor dimensions. As well, IC technology allows the fabrication of arrays of transistors having dimensions of specified ratios capable of being realized with great precision. For instance, the designer can implement an array of bipolar transistors whose emitter-base-junction areas have binary weighted ratios. CMOS technology provides even more flexibility,

with the W and L values of MOS transistors selected to fit a very wide range of device requirements.

3. Resistors, especially those with large values, occupy large silicon areas and should generally be avoided. Instead, constant-current sources implemented with transistors are used in applications such as biasing, or in place of the load resistance of a transistor amplifier.
4. Although small-valued resistors can be fabricated on an IC chip, their absolute values cannot economically be made precise. Rather, arrays of such resistors can be made to have relatively precise ratios.
5. Chip-area considerations also make it impossible to fabricate large-valued capacitors such as those used for signal coupling and bypass in discrete-circuit amplifiers. As a result, IC amplifiers are all direct coupled and use clever techniques to establish signal grounds, as we will see later in this chapter and the next.
6. Small-size capacitors, such as those in the picofarad and fraction-of-a picofarad range, are easy to fabricate in IC MOS technology, with very precise ratios. Capacitors of this kind can be combined with MOS amplifiers and MOS switches to implement a wide variety of signal-processing functions, such as filters (Section 14.8).
7. To pack a large number of devices on the same IC chip, thereby reducing system size and cost while improving reliability, the trend has been to reduce the device dimensions. (For a discussion of Moore's law and device scaling, see Section 5.4.) As of 2019, CMOS process technologies are capable of producing transistors with less than 10-nm channel length. To avoid breaking down the thin oxide layers (less than 1 nm used in these devices), power supplies are limited to 1 V or so. Low power-supply voltages help with another major design challenge: keeping the power dissipated in the chip within acceptable limits. However, the use of such low dc power-supply voltages presents the circuit designer with a host of challenges. For instance, MOS transistors must be operated with overdrive voltages of 0.1 V to 0.2 V. In our study of MOS amplifier design, we will frequently comment on such issues.
8. Currently, the vast majority of analog integrated circuits are designed using CMOS technology. However, BJTs are still used in special analog applications, such as high-quality general-purpose op-amp packages.

SOLID CIRCUITS WITH "FLYING WIRES"

As the importance of transistors grew during the 1950s, packaging became a problem. While it was possible to create smaller and smaller active devices, individual transistor packages had to be large enough to be held in the assembly of an electronic system. To remedy this problem, Texas Instruments (TI) initiated a program of packaging modular electronics to create on a ceramic substrate, larger, more functional system elements. An engineer assinged to the task, Jack Kilby, believed it was necessary to go one step further and design the multiplicity of active and passive elements on a single piece of semiconductor. Thus, in 1958 he created the first "solid circuit," incorporating many transistors and resistors formed on a single slab of germanium and coupled by "flying wire" interconnections to form system elements such as oscillators and amplifiers. In 1959 TI began to use this technique to manufacture the 507 Binary Flip-Flop. While the approach was successful in producing small space-efficient modules, it was not suited for mass production. In 2000 Kilby received the Nobel Prize in Physics for his part in the invention of the integrated circuit.

8.2 IC Biasing: Current Sources and Current Mirrors

Biasing in integrated-circuit design is based on the use of constant-current sources. On an IC chip with a number of amplifier stages, a constant dc current (called a **reference current**) is generated at one location and is then replicated at various other locations for biasing the various amplifier stages through a process known as **current steering**. This approach has the advantage that the effort expended on generating a predictable and stable reference current, usually utilizing a precision resistor external to the chip or a special circuit on the chip, need not be repeated for every amplifier stage. Furthermore, the bias currents of the various stages track each other in case of changes in power-supply voltage or in temperature.

In this section, we study circuit building blocks and techniques used in the bias design of IC amplifiers. These current-source circuits are also used as amplifier load elements, as we will see in Sections 8.3 and 8.4.

8.2.1 The Basic MOSFET Current Source

Figure 8.1 shows the circuit of a simple MOS constant-current source. The heart of the circuit is transistor Q_1, the drain of which is shorted to its gate,[1] thereby forcing it to operate in the saturation mode with

$$I_{D1} = \frac{1}{2}k_n'\left(\frac{W}{L}\right)_1 (V_{GS} - V_{tn})^2 \tag{8.1}$$

where we have neglected channel-length modulation. The drain current of Q_1 is supplied by V_{DD} through resistor R, which in most cases would be outside the IC chip. Since the gate currents are zero,

$$I_{D1} = I_{\text{REF}} = \frac{V_{DD} - V_{GS}}{R} \tag{8.2}$$

where the current through R is considered the reference current of the current source and is denoted I_{REF}. Equations (8.1) and (8.2) can be used together to determine the value required for R to generate a given I_{REF}.

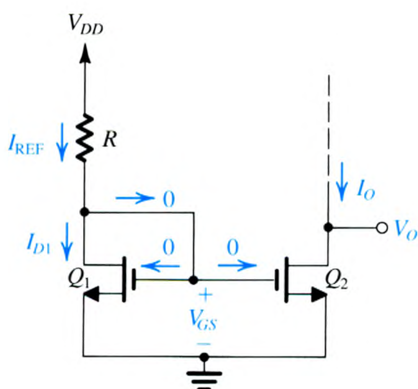


Figure 8.1 Circuit for a basic MOSFET constant-current source. For proper operation, the output terminal, that is, the drain of Q_2, must be connected to a circuit that ensures that Q_2 operates in saturation.

[1] Such a transistor is said to be *diode connected*.

Now consider transistor Q_2: It has the same V_{GS} as Q_1; thus, if we assume that it is operating in saturation, its drain current, which is the output current I_O of the current source, will be

$$I_O = I_{D2} = \frac{1}{2}k_n'\left(\frac{W}{L}\right)_2 (V_{GS}-V_{tn})^2 \tag{8.3}$$

where we have neglected channel-length modulation. Equations (8.1) and (8.3) enable us to relate the output current I_O to the reference current I_{REF} as follows:

$$\frac{I_O}{I_{REF}} = \frac{(W/L)_2}{(W/L)_1} \tag{8.4}$$

This is a simple and attractive relationship: The special connection of Q_1 and Q_2 provides an output current I_O that is related to the reference current I_{REF} by the aspect ratios of the transistors. In other words, the relationship between I_O and I_{REF} is solely determined by the geometries of the transistors. In the special case of identical transistors, $I_O = I_{REF}$, the circuit simply replicates or *mirrors* the reference current in the output terminal. For this reason, the circuit composed of Q_1 and Q_2 is called a **current mirror**. Note that the term is used regardless of the ratio of device dimensions.

8.2.2 The MOS Current Mirror

Figure 8.2 depicts the current-mirror circuit with the input reference current shown as being supplied by a current source for both simplicity and generality. The **current gain** or **current transfer ratio** of the current mirror is given by Eq. (8.4).

We next consider the effect of V_O on I_O. In the description above for the operation of the current source of Fig. 8.1, we assumed Q_2 to be operating in saturation. This is essential if Q_2 is to supply a constant-current output. To ensure that Q_2 is saturated, the circuit to which the drain of Q_2 is to be connected must establish a drain voltage V_O that satisfies the relationship

$$V_O \geq V_{GS} - V_{tn} \tag{8.5}$$

or, equivalently, in terms of the overdrive voltage V_{OV} of Q_1 and Q_2,

$$V_O \geq V_{OV} \tag{8.6}$$

In other words, the current source and the current mirror will operate properly with an output voltage V_O as low as V_{OV}, which is a few tenths of a volt.

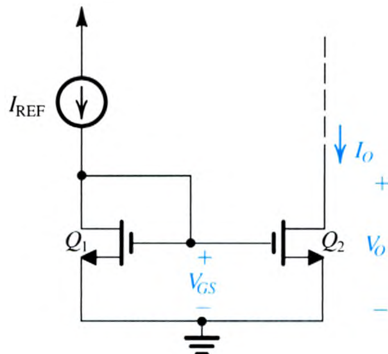


Figure 8.2 Basic MOSFET current mirror.

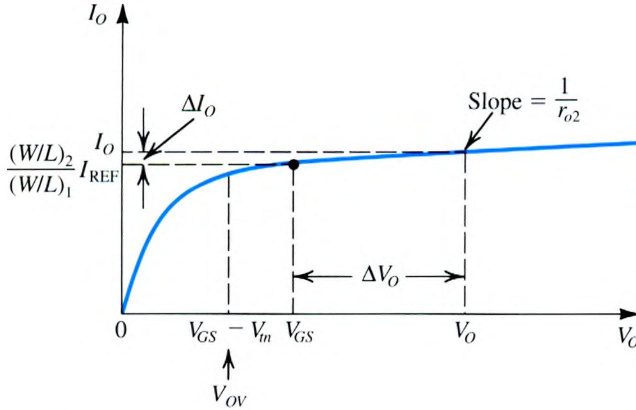


Figure 8.3 Output characteristic of the current source in Fig. 8.1 and the current mirror of Fig. 8.2.

The minimum voltage V_{CSmin} required across a current source or a current mirror for its proper operation is an important characteristic parameter of the current source and the current mirror. Given that for modern CMOS technology, V_{DD} has become as low as one volt or so, it is imperative that V_{CSmin} be as small as possible. For the current source in Fig. 8.1 and the current mirror in Fig. 8.2, V_{CSmin} is equal to V_{OV} of Q_1 and Q_2.

Although thus far neglected, channel-length modulation can have a significant effect on the operation of the current mirror. For the mirror in Fig. 8.2, we see that the output current will be at its nominal value of

$$I_O\,|_{\text{nominal}} = \frac{(W/L)_2}{(W/L)_1} I_{REF} \tag{8.7}$$

at the value of V_O that makes Q_1 and Q_2 have the same V_{DS}; that is, $V_O = V_{GS}$. As V_O deviates from this value, I_O will deviate from the nominal value by ΔI_O,

$$\Delta I_O = \frac{\Delta V_O}{r_{o2}} = \frac{V_O - V_{GS}}{r_{o2}} \tag{8.8}$$

where r_{o2} is the output resistance of Q_2,

$$r_{o2} = \frac{V_{A2}}{I_O\,|_{\text{nominal}}} \tag{8.9}$$

and where V_{A2} is the Early voltage of Q_2. Equations (8.7), (8.8), and (8.9) can be used to find I_O at an arbitrary V_O (that is greater than V_{OV}) as

$$I_O = \frac{(W/L)_2}{(W/L)_1} I_{REF} \left[1 + \frac{V_O - V_{GS}}{V_{A2}}\right] \tag{8.10}$$

This analysis is illustrated graphically in Fig. 8.3, which shows I_O versus V_O. Note that this is the $i_D - v_{DS}$ characteristic curve of Q_2 that corresponds to $v_{GS2} = V_{GS}$. Finally, observe from Eq. (8.9) that the output resistance of the mirror, r_{O2}, is proportional to the Early voltage of Q_2, V_{A2}. Since for a given process technology, V_{A2} is proportional to the channel length of Q_2, to obtain a high output resistance, we normally give the mirror transistor Q_2 a relatively long channel.

Example 8.1

Given $V_{DD} = 3$ V and using $I_{\text{REF}} = 100\ \mu$A, design the circuit of Fig. 8.1 to obtain an output current whose nominal value is 100 μA. Find R if Q_1 and Q_2 are matched and have channel lengths of 1 μm, channel widths of 10 μm, $V_t = 0.7$ V, and $k'_n = 200\ \mu\text{A/V}^2$. What is the lowest possible value of V_O? Assuming that for this process technology, the Early voltage $V'_A = 20$ V/μm, find the output resistance of the current source. Also, find the change in output current resulting from a +1-V change in V_O.

Solution

$$I_{D1} = I_{\text{REF}} = \frac{1}{2}k'_n\left(\frac{W}{L}\right)_1 V_{OV}^2$$

$$100 = \frac{1}{2} \times 200 \times 10\,V_{OV}^2$$

Thus,

$$V_{OV} = 0.316 \text{ V}$$

and

$$V_{GS} = V_t + V_{OV} = 0.7 + .316 \simeq 1 \text{ V}$$

$$R = \frac{V_{DD} - V_{GS}}{I_{\text{REF}}} = \frac{3-1}{0.1 \text{ mA}} = 20 \text{ k}\Omega$$

$$V_{O\min} = V_{OV} \simeq 0.3 \text{ V}$$

Now we can find the output resistance of the current source, r_{o2}. For the transistors used, $L = 1$ μm, thus,

$$V_A = 20 \times 1 = 20 \text{ V}$$

$$r_{o2} = \frac{20 \text{ V}}{100\ \mu\text{A}} = 0.2 \text{ M}\Omega$$

The output current will be 100 μA at $V_O = V_{GS} = 1$ V. If V_O changes by +1 V, the corresponding change in I_O will be

$$\Delta I_O = \frac{\Delta V_O}{r_{o2}} = \frac{1 \text{ V}}{0.2 \text{ M}\Omega} = 5\ \mu\text{A}$$

EXERCISE

D8.1 In the current source of Example 8.1, it is required to reduce the change in output current, ΔI_O, corresponding to a change in output voltage, ΔV_O, of 1 V to 1% of I_O. What should the dimensions of Q_1 and Q_2 be changed to? Assume that Q_1 and Q_2 are to remain matched.
Ans. $L = 5$ μm; $W = 50$ μm

Video Example VE 8.1 Design of a MOS Current Source

For $V_{DD} = 1.3$ V and using $I_{REF} = 20\ \mu A$, it is required to design the circuit of Fig. 8.1 to obtain an output current whose nominal value is $100\ \mu A$. Find R and W_2 if Q_1 and Q_2 have equal channel lengths of $0.5\ \mu m$, $W_1 = 1\ \mu m$, $V_t = 0.4$ V, and $k_n' = 500\ \mu A/V^2$. What is the lowest possible value of V_O? Assuming that for this process technology the Early voltage $V_A' = 5\ V/\mu m$, find the output resistance of the current source. Also, find the change in output current resulting from a +0.2-V change in V_O.

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.


VE 8.1

Related end-of-chapter problem: 8.2

8.2.3 MOS Current-Steering Circuits

As mentioned earlier, once a constant current has been generated, it can be replicated to provide dc bias or load currents for the various amplifier stages in an IC. Current mirrors can obviously be used to implement this current-steering function. Figure 8.4 shows a simple current-steering circuit. Here Q_1 together with R determine the reference current I_{REF}. Transistors Q_1, Q_2, and Q_3 form a two-output current mirror,

$$I_2 = I_{REF}\frac{(W/L)_2}{(W/L)_1} \tag{8.11}$$

$$I_3 = I_{REF}\frac{(W/L)_3}{(W/L)_1} \tag{8.12}$$

To ensure operation in the saturation region, the voltages at the drains of Q_2 and Q_3 are constrained as follows:

$$V_{D2}, V_{D3} \geq -V_{SS} + V_{GS1} - V_{tn}$$

or, equivalently,

$$V_{D2}, V_{D3} \geq -V_{SS} + V_{OV1} \tag{8.13}$$

where V_{OV1} is the overdrive voltage at which Q_1, Q_2, and Q_3 are operating. In other words, the drains of Q_2 and Q_3 will have to remain higher than $-V_{SS}$ by at least the overdrive voltage, which is usually a few tenths of a volt.

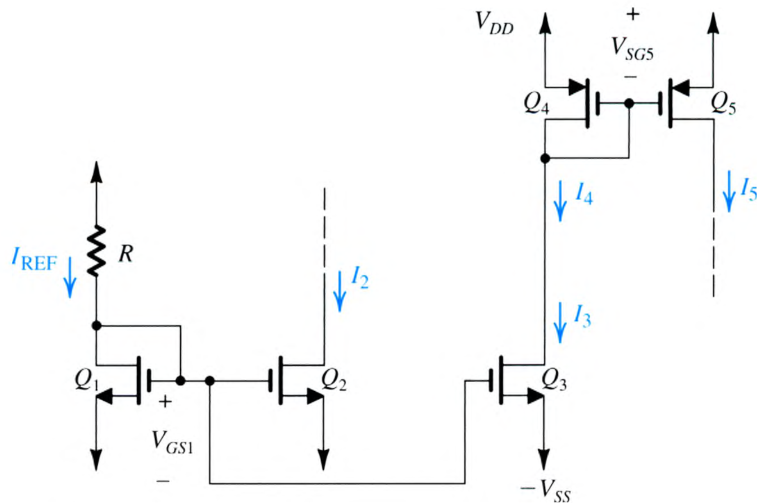


Figure 8.4 A current-steering circuit.

Continuing our discussion of the circuit in Fig. 8.4, we see that current I_3 is fed to the input side of a current mirror formed by PMOS transistors Q_4 and Q_5. This mirror provides

$$I_5 = I_4 \frac{(W/L)_5}{(W/L)_4} \tag{8.14}$$

where $I_4 = I_3$. To keep Q_5 in saturation, its drain voltage should be

$$V_{D5} \leq V_{DD} - |V_{OV5}| \tag{8.15}$$

where V_{OV5} is the overdrive voltage at which Q_5 is operating.

The constant current I_2 generated in the circuit of Fig. 8.4 can be used to bias a source-follower amplifier like the one implemented by transistor Q_6 in Fig. 8.5(a). Similarly, the constant current I_5 can be used as the load for a common-source amplifier like the one implemented with transistor Q_7 in Fig. 8.5(b). We will discuss the use of current sources as load elements for CS amplifiers in Section 8.3.

Finally, an important point to note is that in the circuit of Fig. 8.4, while Q_2 *pulls* its current I_2 from a circuit (not shown in Fig. 8.4), Q_5 *pushes* its current I_5 into a circuit (not shown in Fig. 8.4). Thus Q_5 is appropriately called a **current source**, whereas Q_2 should more properly be called a **current sink**. An IC normally needs both current sources and current sinks. The difference between a current source and a current sink is further illustrated in Fig. 8.6, where $V_{CS\min}$ denotes the minimum voltage needed across the current source (or sink) to operate properly.

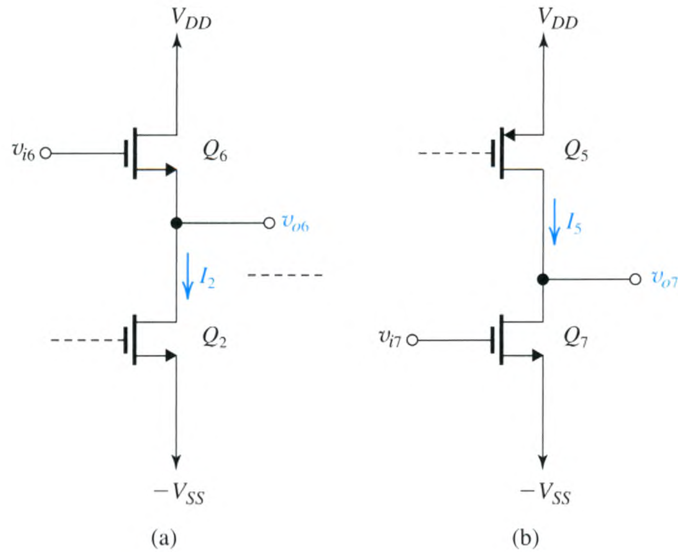


Figure 8.5 Application of the constant currents I_2 and I_5 generated in the current-steering circuit of Fig. 8.4: **(a)** Constant-current I_2 is the bias current for the source follower Q_6, and **(b)** constant-current I_5 is the load current for the common-source amplifier Q_7.

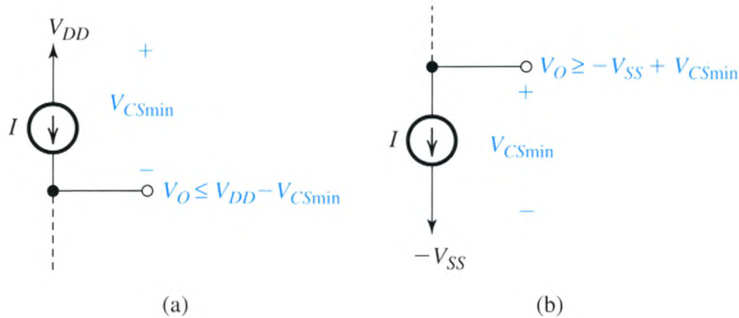


Figure 8.6 (**a**) A current source; and (**b**) a current sink.

EXERCISE

D8.2 For the circuit of Fig. 8.4, let $V_{DD} = V_{SS} = 1.5$ V, $V_{tn} = 0.6$ V, $V_{tp} = -0.6$ V, all channel lengths = 1 μm, $k_n' = 200\ \mu\text{A/V}^2$, $k_p' = 80\ \mu\text{A/V}^2$, and $\lambda = 0$. For $I_{REF} = 10\ \mu$A, find the widths of all transistors to obtain $I_2 = 60\ \mu$A, $I_3 = 20\ \mu$A, and $I_5 = 80\ \mu$A. The voltage at the drain of Q_2 must be allowed to go down to within 0.2 V of the negative supply, and the voltage at the drain of Q_5 must be allowed to go up to within 0.2 V of the positive supply.

Ans. $W_1 = 2.5\ \mu$m; $W_2 = 15\ \mu$m; $W_3 = 5\ \mu$m; $W_4 = 12.5\ \mu$m; $W_5 = 50\ \mu$m

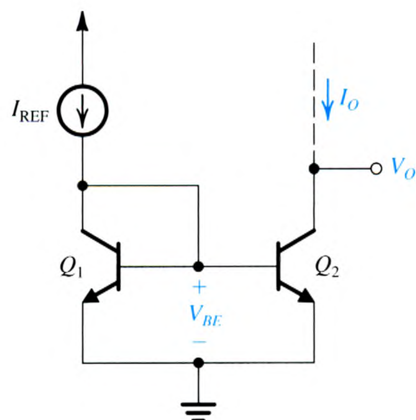


Figure 8.7 The basic BJT current mirror.

8.2.4 BJT Circuits

The basic BJT current mirror is shown in Fig. 8.7. It works in a fashion very similar to that of the MOS mirror. However, there are two important differences: First, the nonzero base current of the BJT (or, equivalently, the finite β) causes an error in the current transfer ratio of the bipolar mirror. Second, the current transfer ratio is determined by the relative areas of the emitter–base junctions of Q_1 and Q_2.

Let's first consider the case of β sufficiently high that we can neglect the base currents. The reference current I_{REF} is passed through the diode-connected transistor Q_1 and thus establishes a corresponding voltage V_{BE}, which in turn is applied between the base and emitter of Q_2. Now, if Q_2 is matched to Q_1 or, more specifically, if the EBJ area of Q_2 is the same as that of Q_1, and thus Q_2 has the same scale current I_S as Q_1, then the collector current of Q_2 will be equal to that of Q_1; that is,

$$I_O = I_{REF}$$

For this to happen, however, Q_2 must be operating in the active mode, which in turn is achieved as long as the collector voltage V_O is 0.3 V or so higher than that of the emitter.

To obtain a current transfer ratio other than unity, say m, we simply arrange that the area of the EBJ of Q_2 is m times that of Q_1. In this case,

$$I_O = mI_{REF} \tag{8.16}$$

In general, the current transfer ratio is given by

$$\frac{I_O}{I_{REF}} = \frac{I_{S2}}{I_{S1}} = \frac{\text{Area of EBJ of } Q_2}{\text{Area of EBJ of } Q_1} \tag{8.17}$$

Alternatively, if the area ratio m is an integer, we can think of Q_2 as equivalent to m transistors, each matched to Q_1 and connected in parallel.

Next we consider the effect of finite transistor β on the current transfer ratio. The analysis for the case in which the current transfer ratio is nominally unity—that is, for the case in which Q_2 is matched to Q_1 —is illustrated in Fig. 8.8. The key point here is that since Q_1 and Q_2 are matched and have the same V_{BE}, their collector currents will be equal. The rest of the analysis is straightforward. A node equation at the collector of Q_1 yields

$$I_{REF} = I_C + 2I_C/\beta = I_C\left(1 + \frac{2}{\beta}\right)$$

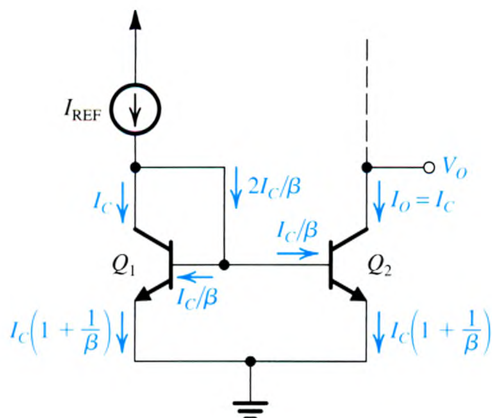


Figure 8.8 Analysis of the current mirror taking into account the finite β of the BJTs.

Finally, since $I_O = I_C$, the current transfer ratio can be found as

$$\frac{I_O}{I_{REF}} = \frac{I_C}{I_C\left(1+\frac{2}{\beta}\right)} = \frac{1}{1+\frac{2}{\beta}} \tag{8.18}$$

Note that as β approaches ∞, I_O/I_{REF} approaches the nominal value of one. For typical values of β, however, the error in the current transfer ratio can be significant. For instance, $\beta = 100$ results in a 2% error in the current transfer ratio. Furthermore, the error due to the finite β increases as the nominal current transfer ratio is increased. We encourage you to show that for a mirror with a nominal current transfer ratio m—that is, one in which $I_{S2} = mI_{S1}$—the actual current transfer ratio is given by

$$\frac{I_O}{I_{REF}} = \frac{m}{1+\frac{m+1}{\beta}} \tag{8.19}$$

In common with the MOS current mirror, the BJT mirror has a finite output resistance R_o,

$$R_o \equiv \frac{\Delta V_O}{\Delta I_O} = r_{o2} = \frac{V_{A2}}{I_O} \tag{8.20}$$

where V_{A2} and r_{o2} are the Early voltage and the output resistance, respectively, of Q_2. Thus, even if we neglect the error due to finite β, the output current I_O will be at its nominal value only when Q_2 has the same V_{CE} as Q_1, namely, at $V_O = V_{BE}$. As V_O is changed, I_O will correspondingly change. Taking both the finite β and the finite R_o into account, we can express the output current of a BJT mirror with a nominal current transfer ratio m as

$$I_O = I_{REF}\frac{m}{1+\frac{m+1}{\beta}}\left(1+\frac{V_O - V_{BE}}{V_{A2}}\right) \tag{8.21}$$

where we note that the error term due to the Early effect is expressed in a form that shows that it reduces to zero for $V_O = V_{BE}$.

EXERCISE

8.3 Consider a BJT current mirror with a nominal current transfer ratio of unity. Let the transistors have $I_S = 10^{-15}$ A, $\beta = 100$, and $V_A = 100$ V. For $I_{REF} = 1$ mA, find I_O when $V_O = 5$ V. Also, find the output resistance.
Ans. 1.02 mA; 100 kΩ

A Simple Current Source As in the MOS case, the basic BJT current mirror can be used to implement a simple current source, as shown in Fig. 8.9. Here the reference current is

$$I_{REF} = \frac{V_{CC} - V_{BE}}{R} \tag{8.22}$$

where V_{BE} is the base–emitter voltage corresponding to the desired value of I_{REF}. For matched transistors, the output current I_O is given by

$$I_O = \frac{I_{REF}}{1 + (2/\beta)}\left(1 + \frac{V_O - V_{BE}}{V_A}\right) \tag{8.23}$$

The output resistance of this current source is r_o of Q_2,

$$R_o = r_{o2} \simeq \frac{V_A}{I_O} \simeq \frac{V_A}{I_{REF}} \tag{8.24}$$

EXERCISE

D8.4 Assuming the availability of BJTs with scale currents $I_S = 10^{-15}$ A, $\beta = 100$, and $V_A = 50$ V, design the current-source circuit of Fig. 8.9 to provide an output current $I_O = 0.5$ mA at $V_O = 2$ V. The power supply $V_{CC} = 5$ V. Give the values of I_{REF}, R, and $V_{O\min}$. Also, find I_O at $V_O = 5$ V.
Ans. 0.497 mA; 8.71 kΩ; 0.3 V; 0.53 mA

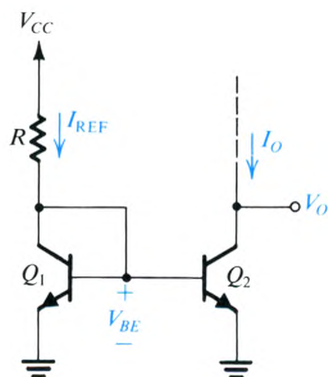


Figure 8.9 A simple BJT current source.

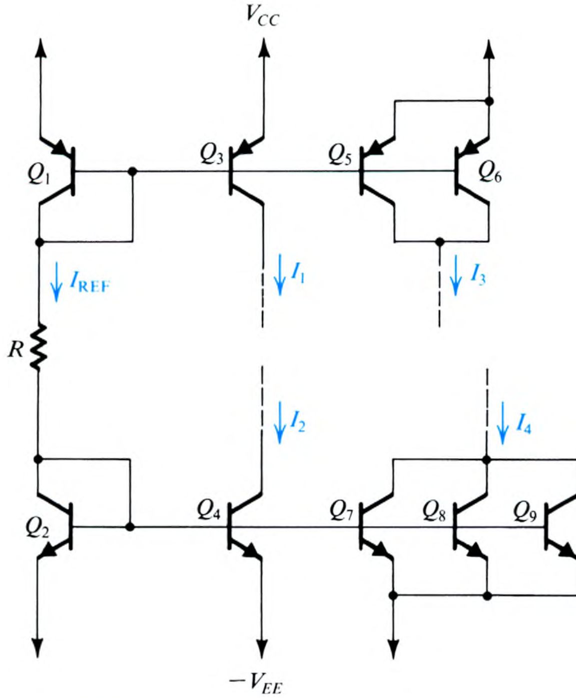


Figure 8.10 Generation of a number of constant currents of various magnitudes.

Current Steering To generate bias currents for different amplifier stages in an IC, the current-steering approach described for MOS circuits can be applied in the bipolar case. As an example, consider the circuit shown in Fig. 8.10. The dc reference current I_{REF} is generated in the branch that consists of the diode-connected transistor Q_1, resistor R, and the diode-connected transistor Q_2:

$$I_{REF} = \frac{V_{CC} + V_{EE} - V_{EB1} - V_{BE2}}{R} \tag{8.25}$$

Now, for simplicity, assume that all the transistors have high β and thus that the base currents are negligibly small. We will also neglect the Early effect. The diode-connected transistor Q_1 forms a current mirror with Q_3; thus Q_3 will supply a constant current I_1 equal to I_{REF}. Transistor Q_3 can supply this current to any load as long as the voltage that develops at the collector does not exceed ($V_{CC} - 0.3$ V); otherwise Q_3 would enter the saturation region.

To generate a dc current twice the value of I_{REF}, two transistors, Q_5 and Q_6, each of them matched to Q_1, are connected in parallel. The combination forms a mirror with Q_1, so that $I_3 = 2I_{REF}$. Note that the parallel combination of Q_5 and Q_6 is equivalent to a transistor with an EBJ area double that of Q_1, which is precisely what is done when this circuit is fabricated in IC form.

Transistor Q_4 forms a mirror with Q_2; thus Q_4 provides a constant current I_2 equal to I_{REF}. Note that while Q_3 *sources* its current to parts of the circuit whose voltage should not exceed ($V_{CC} - 0.3$ V), Q_4 *sinks* its current from parts of the circuit whose voltage should not decrease below ($-V_{EE} + 0.3$ V). Finally, to generate a current three times I_{REF}, three transistors, Q_7, Q_8, and Q_9, each of which is matched to Q_2, are connected in parallel, and the combination

is placed in a mirror configuration with Q_2. Again, in an IC implementation, Q_7, Q_8, and Q_9 would be replaced with a transistor having a junction area three times that of Q_2.

Video Example VE 8.2 A BJT Current Steering Circuit

Find the voltages at all nodes and the currents through all branches in the circuit of Fig. VE8.2. Assume $|V_{BE}| = 0.7$ V and $\beta = \infty$.

Figure VE8.2 Circuit for Video Example 8.2.

VE 8.2

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 8.15

EXERCISE

8.5 Figure E8.5 shows an N-output current mirror. Assuming that all transistors are matched and have finite β and ignoring the effect of finite output resistances, show that

$$I_1 = I_2 = \cdots = I_N = \frac{I_{REF}}{1 + (N+1)/\beta}$$

For $\beta = 100$, find the maximum number of outputs for an error not exceeding 10%.

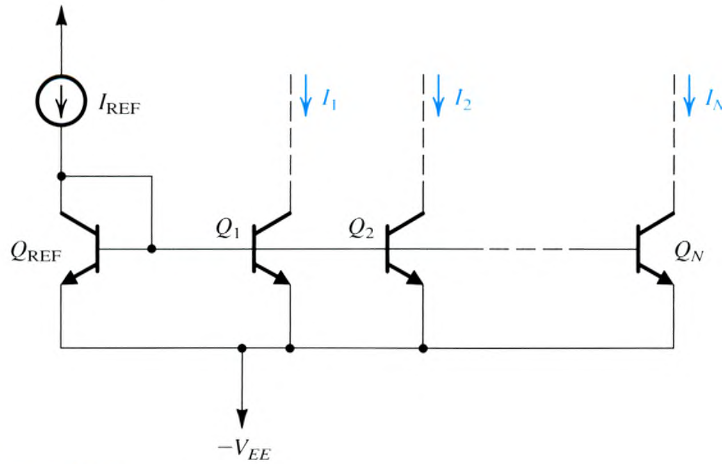


Figure E8.5

Ans. 9

A Bipolar Mirror with Base-Current Compensation Figure 8.11 shows a bipolar current mirror with a current transfer ratio that is much less dependent on β than that of the simple current mirror. The reduced dependence on β is achieved by including transistor Q_3, whose emitter supplies the base currents of Q_1 and Q_2. The sum of the base currents is then divided by $(\beta_3 + 1)$, resulting in a much smaller error current that has to be supplied by I_{REF}. Detailed analysis is shown on the circuit diagram; it is based on the assumption that Q_1 and Q_2 are matched and thus have equal collector currents, I_C. A node equation at the node labeled x gives

$$I_{REF} = I_C\left[1 + \frac{2}{\beta(\beta+1)}\right]$$

Since

$$I_O = I_C$$

the current transfer ratio of the mirror will be

$$\frac{I_O}{I_{REF}} = \frac{1}{1 + 2/(\beta^2 + \beta)}$$
$$\simeq \frac{1}{1 + 2/\beta^2} \tag{8.26}$$

which means that the error due to finite β has been reduced from $2/\beta$ in the simple mirror to $2/\beta^2$, a tremendous improvement. Unfortunately, however, the output resistance remains

Figure 8.11 A current mirror with base-current compensation.

approximately equal to that of the simple mirror, namely r_o. Finally, note that if a reference current I_{REF} is not available, we simply connect node x to the power supply, V_{CC}, through a resistance R. The result is a reference current given by

$$I_{REF} = \frac{V_{CC} - V_{BE1} - V_{BE3}}{R} \tag{8.27}$$

8.2.5 Small-Signal Operation of Current Mirrors

In addition to their use in biasing, current mirrors are sometimes used as current amplifiers. It is therefore useful to derive the small-signal parameters of the current mirror, that is, R_{in}, A_{is}, and R_o. Here, A_{is} is the short-circuit current gain.

Figure 8.12(a) shows a MOS current mirror biased with a dc input current I_{D1} and fed with a small-signal input current i_i. Note that V_{GS} and I_{D2} are the resulting dc quantities, while v_{gs} and i_o are signal quantities. Although we are not showing the circuit to which the output terminal is connected, we are assuming that the voltage at the drain of Q_2 exceeds the minimum required to keep Q_2 in saturation. Figure 8.12(b) shows the small-signal equivalent circuit model of the current mirror. In the following, we determine the value of the model parameters.

Replacing Q_1 and Q_2 with their small-signal models results in the circuit in Fig. 8.12(c). Notice that the controlled current source $g_{m1}v_{gs}$ appears across its control voltage v_{gs} and thus can be replaced with a resistance, $1/g_{m1}$, as shown in Fig. 8.12(d). For the latter circuit we can obtain

$$R_{in} = r_{o1} \left\| \frac{1}{g_{m1}} \simeq \frac{1}{g_{m1}} \right. \tag{8.28}$$

$$R_o = r_{o2} \tag{8.29}$$

$$A_{is} \equiv \left. \frac{i_o}{i_i} \right|_{v_{d2}=0} = \frac{g_{m2}\,v_{gs}}{i_i}$$

where

$$v_{gs} = i_i R_{in} = \frac{i_i}{g_{m_1} + \dfrac{1}{r_{o1}}}$$

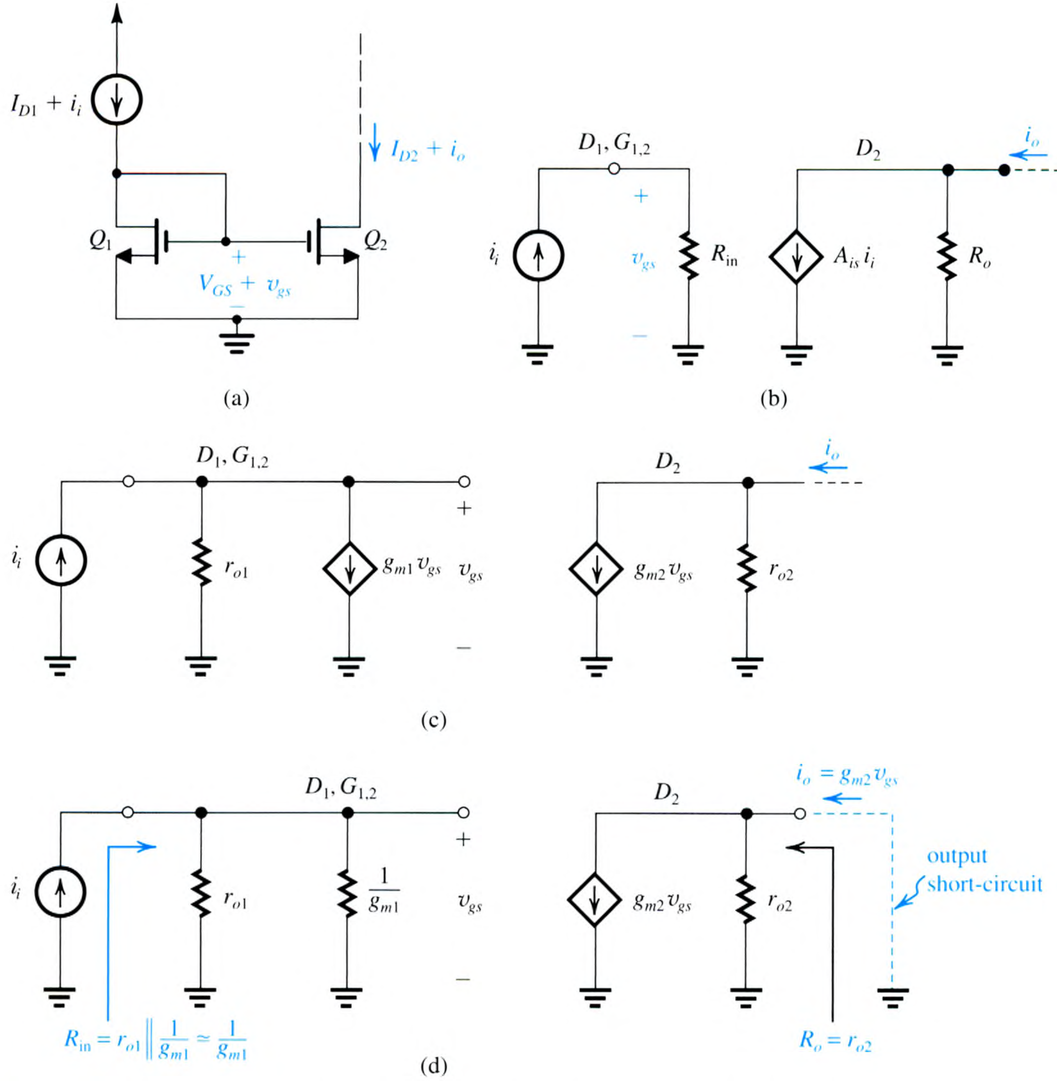


Figure 8.12 Obtaining the small-signal parameters of the MOS current mirror as a current amplifier.

Thus,

$$A_{is} = \frac{g_{m2}}{g_{m1}} \frac{1}{1+(1/g_{m1}r_{o1})} \tag{8.30}$$

Substituting for $g_{m1,2} = \mu_n C_{ox}(W/L)_{1,2}V_{OV}$, where V_{OV} is the overdrive voltage at which Q_1 and Q_2 are operating, yields for the short-circuit current gain

$$A_{is} = \frac{(W/L)_2}{(W/L)_1}\left[\frac{1}{1+(1/g_{m1}r_{o1})}\right] \tag{8.31}$$

which is equal to the dc or large-signal current transfer ratio except for a small error factor represented by the quantity in the square brackets.

We conclude that the current mirror is an excellent current amplifier: It has a relatively low input resistance ($1/g_{m1}$), a relatively high output resistance (r_{o2}), and a gain determined by the aspect ratios of the MOSFETs. We can use a similar procedure to find the small-signal parameters of the bipolar mirror (Problem 8.25).

EXERCISE

D8.6 The MOSFETs in the current mirror of Fig. 8.12(a) have equal channel lengths, $\mu_n C_{ox} = 400\ \mu\text{A/V}^2$, and $V'_A = 20\ \text{V}/\mu\text{m}$. If the input bias current is 100 μA, find W_1, W_2, L_1, and L_2 to obtain a short-circuit current gain that is ideally 5 A/A, an input resistance of 1 kΩ, and an output resistance of 40 kΩ. What is the expected percentage error in the current transfer ratio?
Ans. 12.5 μm; 62.5 μm; 1 μm; 1 μm; and −0.5%

THE INTEGRATED CIRCUIT

In 1959, at the same time that Kilby and TI applied for a patent on "miniaturized electronic circuits," Robert Noyce (a cofounder of Fairchild Semiconductor and later of Intel) filed a patent on the "monolithic silicon-based integrated circuit." He later acknowledged the critical importance of Kurt Lehovac's idea of using reverse-biased junctions to isolate multiple devices on a single die. Lehovac, of Sprague Electric Company, also filed a patent in 1959. Regrettably, Noyce, who died in 1990, did not live to share in the Nobel Prize with Kilby.

8.3 The Basic Gain Cell

8.3.1 The CS and CE Amplifiers with Current-Source Loads

The basic gain cell in an IC amplifier is a common-source (CS) or common-emitter (CE) transistor loaded with a constant-current source, as shown in Figs. 8.13(a) and (b). These circuits are similar to the CS and CE amplifiers studied in Section 7.3, except that here we have replaced the resistances R_D and R_C with constant-current sources. We have done so for three reasons:

1. The constant-current source doubles as both the biasing and the load device.
2. From Eq. (7.18), the maximum gain obtained from a CS amplifier with a resistive load is approximately $2V_{DD}/V_{OV}$. Since for modern CMOS process technologies $V_{DD} \simeq 1$ V and $V_{OV} \simeq 0.2$ V, the gain is limited to about 10 V/V. Higher voltage gains can be achieved by replacing R_D with a current source having a large output resistance R_o. This will be demonstrated in this and the next sections.
3. As we mentioned in Section 8.1, in IC technology, using current sources implemented with transistors is easier and more economical in terms of chip area than using resistances with reasonably precise values.

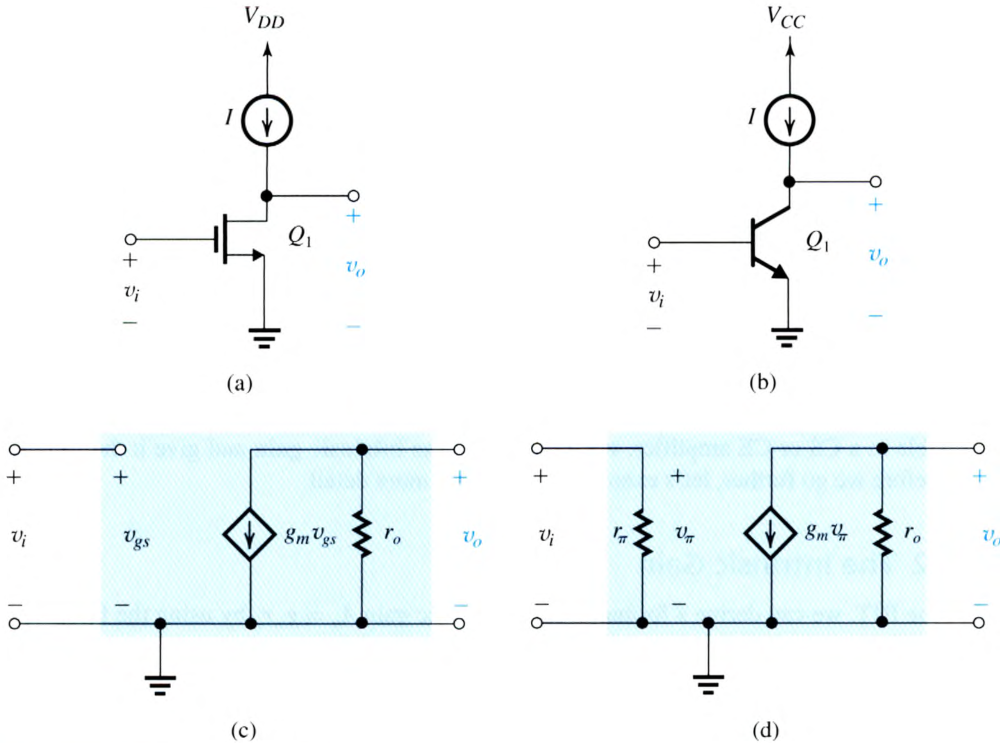


Figure 8.13 The basic gain cells of IC amplifiers: (**a**) current-source- or active-loaded common-source amplifier; (**b**) current-source- or active-loaded common-emitter amplifier; (**c**) small-signal equivalent circuit of (**a**); and (**d**) small-signal equivalent circuit of (**b**).

The circuits in Figs. 8.13(a) and (b) are said to be **current-source loaded** or **active loaded**. Before we consider the small-signal analysis of the active-loaded CS and CE amplifiers, a word on their dc bias is in order. We know that in each circuit, Q_1 is biased at $I_D = I$ and $I_C = I$. But what determines the dc voltages at the drain (collector) and at the gate (base)? Usually, these gain cells will be part of larger circuits in which negative feedback is used to fix the values of V_{DS} and V_{GS} (V_{CE} and V_{BE}). In the next chapter we will begin to see complete IC amplifiers including biasing. For the time being, let's assume that the MOS transistor in Fig. 8.13(a) is biased to operate in the saturation region and that the BJT in Fig. 8.13(b) is biased to operate in the active region. We will often refer to both the MOSFET and the BJT as operating in the "active region."

Small-signal analysis of the current-source-loaded CS and CE amplifiers can be performed using their equivalent-circuit models, shown respectively in Fig. 8.13(c) and (d). Notice that since we assume the current-source load to be ideal, it is represented in the models by an infinite resistance. Practical current sources have finite output resistance, as we have seen in the previous section. For now note that the CS and CE amplifiers of Fig. 8.13 are in effect operating in an open-circuit fashion. The only resistance between their output node and ground is the output resistance of the transistor itself, r_o. Thus the voltage gain obtained in these circuits is the *maximum possible* for a CS or a CE amplifier.

From Fig. 8.13(c) we obtain for the active-loaded CS amplifier:

$$R_{\text{in}} = \infty \tag{8.32}$$

$$A_{vo} = -g_m r_o \tag{8.33}$$

$$R_o = r_o \tag{8.34}$$

Similarly, from Fig. 8.13(d) we obtain for the active-loaded CE amplifier:

$$R_{\text{in}} = r_\pi \tag{8.35}$$

$$A_{vo} = -g_m r_o \tag{8.36}$$

$$R_o = r_o \tag{8.37}$$

Thus both circuits realize a voltage gain of magnitude $g_m r_o$. Since this is the maximum gain possible in a CS or CE amplifier, we refer to it as the **intrinsic gain** and give it the symbol A_0. Before we go further, let's examine A_0 in a little more detail.

8.3.2 The Intrinsic Gain

For the BJT, we can derive a formula for the intrinsic gain $A_0 = g_m r_o$ by using the following formulas for g_m and r_o:

$$g_m = \frac{I_C}{V_T} \tag{8.38}$$

$$r_o = \frac{V_A}{I_C} \tag{8.39}$$

The result is

$$A_0 = g_m r_o = \frac{V_A}{V_T} \tag{8.40}$$

Thus A_0 is simply the ratio of the Early voltage V_A, which is a technology-determined parameter, and the thermal voltage V_T, which is a physical parameter (approximately 0.025 V at room temperature). The value of V_A ranges from 5 V to 35 V for modern IC fabrication processes to 100 V to 130 V for the older, so-called high-voltage processes (see Appendix G). As a result, the value of A_0 will be in the range of 200 V/V to 5000 V/V, with the lower values characteristic of modern small-feature-size devices. It is important to note that for a given bipolar-transistor fabrication process, A_0 is independent of the transistor junction area and of its bias current. This is not the case for the MOSFET, as we shall now see.

Recall from our study of the MOSFET g_m in Section 7.2 that there are three possible expressions for g_m. Two of these are particularly useful for our purposes here:

$$g_m = \frac{I_D}{V_{OV}/2} \tag{8.41}$$

$$g_m = \sqrt{2\mu_n C_{ox}(W/L)}\sqrt{I_D} \tag{8.42}$$

For the MOSFET r_o we have

$$r_o = \frac{V_A}{I_D} = \frac{V_A' L}{I_D} \tag{8.43}$$

where V_A is the Early voltage and V_A' is the technology-dependent component of the Early voltage. Utilizing each of the g_m expressions together with the expression for r_o, we obtain for A_0,

$$A_0 = \frac{V_A}{V_{OV}/2} \tag{8.44}$$

which can be expressed in the alternate forms

$$A_0 = \frac{2V_A'L}{V_{OV}} \tag{8.45}$$

and

$$A_0 = \frac{V_A'\sqrt{2(\mu_n C_{ox})(WL)}}{\sqrt{I_D}} \tag{8.46}$$

The expression in Eq. (8.44) is the one most directly comparable to that of the BJT (Eq. 8.40). Here, however, we note the following:

1. The quantity in the denominator is $V_{OV}/2$, which is a design parameter. Although the value of V_{OV} that designers use for modern submicron technologies has been steadily decreasing, it is still about 0.15 V to 0.3 V. Thus $V_{OV}/2$ is 0.075 V to 0.15 V, which is 3 to 6 times higher than V_T. Furthermore, there are reasons for selecting higher values for V_{OV} (to be discussed in later chapters).
2. The numerator quantity is both process dependent (through V_A') and device dependent (through L), and its value has been steadily decreasing with the scaling down of the technology (see Appendix K).
3. From Eq. (8.45) we see that for a given technology (i.e., a given value of V_A') the intrinsic gain A_0 can be increased by using a longer MOSFET and operating it at a lower V_{OV}. As usual, however, there are design trade-offs. For instance, we will see in Chapter 10 that increasing L and lowering V_{OV} result, independently, in decreasing the amplifier bandwidth.

As a result, the intrinsic gain realized in a MOSFET fabricated in a modern short-channel technology is only 10 V/V to 40 V/V, an order of magnitude lower than that for a BJT.

The alternative expression for the MOSFET A_0 given in Eq. (8.46) reveals a very interesting fact: For a given process technology (V_A' and $\mu_n C_{ox}$) and a given device (W and L), the intrinsic gain is inversely proportional to $\sqrt{I_D}$. This is illustrated in Fig. 8.14, which shows a typical plot for A_0 versus the bias current I_D. The plot confirms that the gain increases as the bias current is lowered. The gain, however, levels off at very low currents. This is because the MOSFET enters the **subthreshold region** of operation (Section 5.4.2), where it becomes very much like a BJT with an exponential current–voltage characteristic. The intrinsic gain then becomes constant, just like that of a BJT. Note, however, that although higher gain is obtained at lower values of I_D, the price paid is a lower g_m (Eq. 8.42) and less ability to drive capacitive loads, and thus a decrease in bandwidth. This point will be studied in Chapter 10.

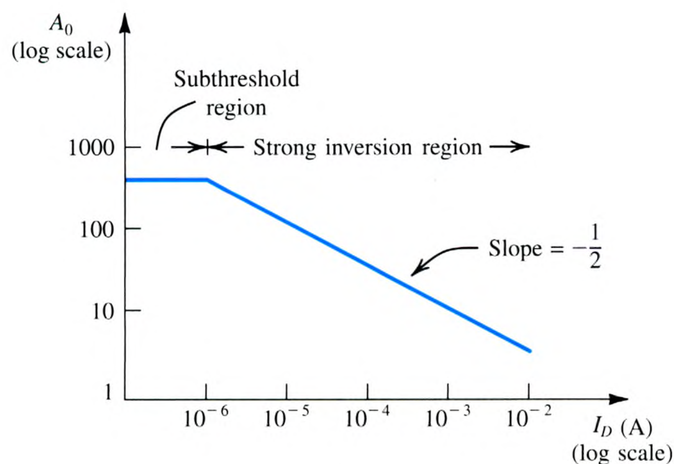


Figure 8.14 The intrinsic gain of the MOSFET versus bias current I_D. Outside the subthreshold region, this is a plot of $A_0 = V_A'\sqrt{2\mu_n C_{ox} WL/I_D}$ for the case: $\mu_n C_{ox} = 20\ \mu\text{A/V}^2$, $V_A' = 20\ \text{V}/\mu\text{m}$, $L = 2\ \mu\text{m}$, and $W = 20\ \mu\text{m}$.

Example 8.2

We wish to compare the values of g_m, R_{in}, R_o, and A_0 for a CS amplifier that is designed using an NMOS transistor with $L = 0.4\ \mu\text{m}$ and $W = 4\ \mu\text{m}$ and fabricated in a 0.25-μm technology specified to have $\mu_n C_{ox} = 267\ \mu\text{A/V}^2$ and $V_A' = 10\ \text{V}/\mu\text{m}$, with those for a CE amplifier designed using a BJT fabricated in a process with $\beta = 100$ and $V_A = 10$ V. Assume that both devices are operating at a drain (collector) current of 100 μA.

Solution

For simplicity, we shall neglect the Early effect in the MOSFET in determining V_{OV}; thus,

$$I_D = \frac{1}{2}(\mu_n C_{ox})\left(\frac{W}{L}\right)V_{OV}^2$$

$$100 = \frac{1}{2} \times 267 \times \left(\frac{4}{0.4}\right)V_{OV}^2$$

resulting in

$$V_{OV} = 0.27\ \text{V}$$

$$g_m = \frac{2I_D}{V_{OV}} = \frac{2 \times 0.1}{0.27} = 0.74\ \text{mA/V}$$

$$R_{in} = \infty$$

$$r_o = \frac{V'_A L}{I_D} = \frac{10 \times 0.4}{0.1} = 40 \text{ k}\Omega$$
$$R_o = r_o = 40 \text{ k}\Omega$$
$$A_0 = g_m r_o = 0.74 \times 40 = 29.6 \text{ V/V}$$

For the CE amplifier we have

$$g_m = \frac{I_C}{V_T} = \frac{0.1 \text{ mA}}{0.025 \text{ V}} = 4 \text{ mA/V}$$
$$R_{\text{in}} = r_\pi = \frac{\beta}{g_m} = \frac{100}{4} = 25 \text{ k}\Omega$$
$$r_o = \frac{V_A}{I_C} = \frac{10}{0.1} = 100 \text{ k}\Omega$$
$$R_o = r_o = 100 \text{ k}\Omega$$
$$A_0 = g_m r_o = 4 \times 100 = 400 \text{ V/V}$$

EXERCISE

8.7 A CS amplifier utilizes an NMOS transistor with $L = 0.36\ \mu\text{m}$ and $W/L = 10$; it was fabricated in a 0.18-μm CMOS process for which $\mu_n C_{ox} = 387\ \mu\text{A/V}^2$ and $V'_A = 5\ \text{V}/\mu\text{m}$. Find the values of g_m and A_0 obtained at $I_D = 10\ \mu\text{A}$, $100\ \mu\text{A}$, and 1 mA.
Ans. 0.28 mA/V, 50 V/V; 0.88 mA/V, 15.8 V/V; 2.78 mA/V, 5 V/V

8.3.3 Effect of the Output Resistance of the Current-Source Load

The current-source load of the CS amplifier in Fig. 8.13(a) can be implemented using a PMOS transistor biased in the saturation region to provide the required current I, as shown in Fig. 8.15(a). We can use the large-signal MOSFET model (Section 5.2, Fig. 5.18) to model Q_2 as shown in Fig. 8.15(b), where

$$I = \frac{1}{2}(\mu_p C_{ox})\left(\frac{W}{L}\right)_2 \left[V_{DD} - V_G - |V_{tp}|\right]^2 \tag{8.47}$$

and

$$r_{o2} = \frac{|V_{A2}|}{I} \tag{8.48}$$

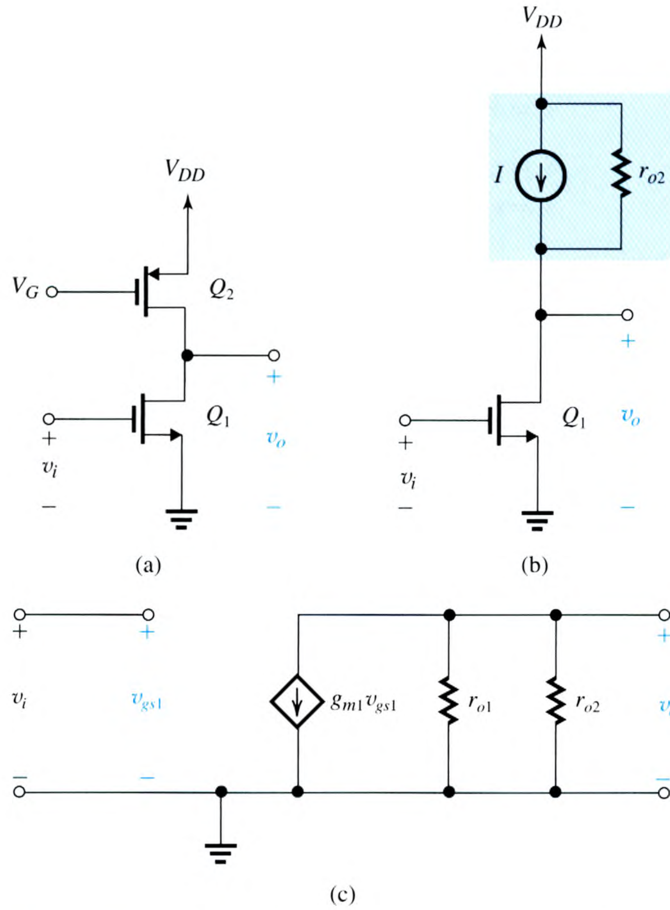


Figure 8.15 (**a**) The CS amplifier with the current-source load implemented with a p-channel MOSFET Q_2; (**b**) the circuit with Q_2 replaced with its large-signal model; (**c**) small-signal equivalent circuit of the amplifier.

Thus the current-source load no longer has an infinite resistance; instead, it has a finite output resistance r_{o2}. This resistance will in effect appear in parallel with r_{o1}, as shown in the amplifier equivalent-circuit model in Fig. 8.15(c), from which we obtain

$$A_v \equiv \frac{v_o}{v_i} = -g_{m1}(r_{o1} \parallel r_{o2}) \tag{8.49}$$

Not surprisingly, the finite output resistance of the current-source load reduces the magnitude of the voltage gain from $(g_{m1}r_{o1})$ to $g_{m1}(r_{o1} \parallel r_{o2})$. This reduction can be substantial. For instance, if Q_2 has an Early voltage equal to that of Q_1, $r_{o2} = r_{o1}$ and the gain is reduced by half,

$$A_v = -\frac{1}{2}g_m r_o \tag{8.50}$$

Finally, we note that a similar development can be used for the bipolar case.

Example 8.3

A practical circuit implementation of the common-source amplifier with a current-source load is shown in Fig. 8.16(a). Here the current-source transistor Q_2 is the output transistor of a current mirror formed by Q_2 and Q_3 and fed with a reference current I_{REF}. The NMOS version of this current source was studied in Section 8.1. Assume that Q_2 and Q_3 are matched.

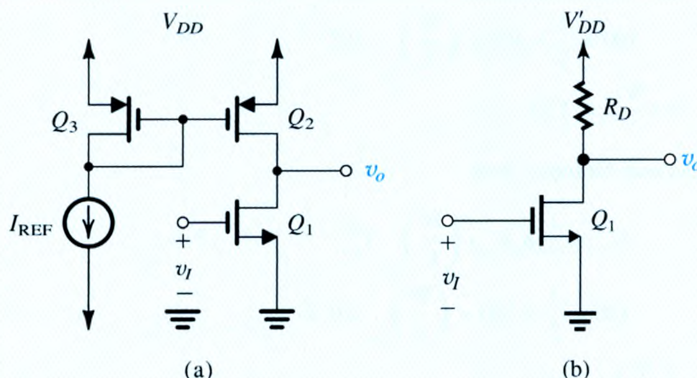


Figure 8.16 (**a**) Practical implementation of the common-source amplifier with a current-source load. (**b**) Replacing the current-source load with a resistance R_D connected to a new power-supply voltage V'_{DD}.

Let $V_{DD} = 1.8$ V, $V_{tn} = -V_{tp} = 0.5$ V, $\mu_n C_{ox} = 4\,\mu_p C_{ox} = 400$ μA/V^2, $|V_A| = 5$ V for all transistors, and $I_{REF} = 100$ μA.

(a) Find the dc component of v_I and the W/L ratios so that all transistors operate at $|V_{OV}| = 0.2$ V.
(b) Determine the small-signal voltage gain.
(c) What is the allowable range of signal swing at the output for almost-linear operation?
(d) If the current-source load is replaced with a resistance R_D connected to a power supply V'_{DD} as shown in Fig. 8.16(b), find the value of R_D and V'_{DD} to keep I_D, the voltage gain, and the output signal swing unchanged.

Solution

(a) To operate Q_1 and $V_{OV} = 0.2$ V, V_{GS1} must be

$$V_{GS1} = V_{tn} + V_{OV} = 0.5 + 0.2 = 0.7\text{ V}$$

and thus the dc component of v_I must be

$$V_I = V_{GS1} = 0.7\text{ V}$$

Example 8.3 *continued*

Next, we determine $(W/L)_1$ from

$$I_{D1} = \frac{1}{2}(\mu_n C_{ox})\left(\frac{W}{L}\right)_1 V_{OV}^2$$

Substituting $I_{D1} = I_{REF} = 100\ \mu A$,

$$100 = \frac{1}{2} \times 400 \times \left(\frac{W}{L}\right)_1 \times 0.2^2$$

$$\Rightarrow (\frac{W}{L})_1 = 12.5$$

The W/L ratios for Q_2 and Q_3 can now be found from

$$I_{D2,3} = \frac{1}{2}(\mu_p C_{ox})\left(\frac{W}{L}\right)_{2,3} V_{OV}^2$$

$$100 = \frac{1}{2} \times 100 \times \left(\frac{W}{L}\right)_{2,3} \times 0.2^2$$

$$\Rightarrow \left(\frac{W}{L}\right)_{2,3} = 50$$

(b) To obtain the small-voltage signal voltage gain, we first determine the small-signal parameters g_{m1}, r_{o1}, and r_{o2} as follows:

$$g_{m1} = \frac{2I_{D1}}{V_{OV1}} = \frac{2 \times 0.1\ \text{mA}}{0.2\ \text{V}} = 1\ \text{mA/V}$$

$$r_{o1} = \frac{V_{A1}}{I_{D1}} = \frac{5\ \text{V}}{0.1\ \text{mA}} = 50\ \text{k}\Omega$$

$$r_{o2} = \frac{|V_{A2}|}{I_{D2}} = \frac{5\ \text{V}}{0.1\ \text{mA}} = 50\ \text{k}\Omega$$

The voltage gain A_v can now be found as

$$\begin{aligned} A_v &= -g_{m1}(r_{o1} \| r_{o2}) \\ &= -1(50\|50) \\ &= -25\ \text{V/V} \end{aligned}$$

(c) The upper limit of the output signal swing is determined by Q_2 leaving the saturation region. This will occur if v_O reaches $V_{DD} - |V_{OV2}|$, thus

$$v_{O\max} = 1.8 - 0.2 = 1.6\ \text{V}$$

The lower limit of the output signal swing is determined by Q_1 leaving the saturation region. This will occur if v_O falls below V_{OV1}, thus

$$v_{O\min} = 0.2\ \text{V}$$

Thus, the range of linear signal swing at the output is given by

$$0.2\ \text{V} \leq v_O \leq 1.6\ \text{V}$$

(d) If the current-source load is replaced with a resistance R_D as in Fig. 8.16(b), to keep the gain unchanged,

$$R_D = r_{o2} = 50\ \text{k}\Omega$$

To keep the output signal swing unchanged, we bias Q_1 so that V_{DS1} is equal to the mid point of the output swing, that is,

$$V_{DS1} = 0.9\ \text{V}$$

Now, for I_D to be the same as before, that is, $I_D = 0.1$ mA, the new supply voltage must be,

$$V_{DD}' = V_{DS1} + I_D R_D$$
$$= 0.9 + 0.1 \times 50 = 5.9\ \text{V}$$

which is much larger than the supply voltage possible with this 0.18-μm CMOS technology (1.8 V). Needless to say, fabricating a 50 kΩ resistance with precise value on the IC is an expensive endeavor. This illustrates the need for using current-source loads.

EXERCISES

8.8 A CMOS common-source amplifier such as that in Fig. 8.16(a), fabricated in a 0.18-μm technology, has $W/L = 7.2\ \mu\text{m}/0.36\ \mu\text{m}$ for all transistors, $k_n' = 387\ \mu\text{A/V}^2$, $k_p' = 86\ \mu\text{A/V}^2$, $I_{\text{REF}} = 100\ \mu\text{A}$, $V_{An}' = 5\ \text{V}/\mu\text{m}$, and $|V_{Ap}'| = 6\ \text{V}/\mu\text{m}$. Find g_{m1}, r_{o1}, r_{o2}, and the voltage gain.
Ans. 1.24 mA/V; 18 kΩ; 21.6 kΩ; –12.2 V/V

8.9 Consider the active-loaded CE amplifier when the constant-current source I is implemented with a *pnp* transistor. Let $I = 0.1$ mA, $|V_A| = 50$ V (for both the *npn* and the *pnp* transistors), and $\beta = 100$. Find R_{in}, r_o (for each transistor), g_m, A_0, and the amplifier voltage gain.
Ans. 25 kΩ; 0.5 MΩ; 4 mA/V; 2000 V/V; –1000 V/V

8.3.4 Increasing the Gain of the Basic Cell

We conclude this section by considering a question: How can we increase the voltage gain obtained from the basic gain cell? The answer lies in finding a way to raise the level of the output resistance of both the amplifying transistor and the load transistor. To increase the output resistance of the amplifying transistor, we need a circuit that passes the current $g_m v_i$ provided by the amplifying transistor right through, but increases the resistance from r_o to a much larger value. This requirement is illustrated in Fig. 8.17. Figure 8.17(a) shows the CS amplifying transistor Q_1 together with its output equivalent circuit. Note that for the time

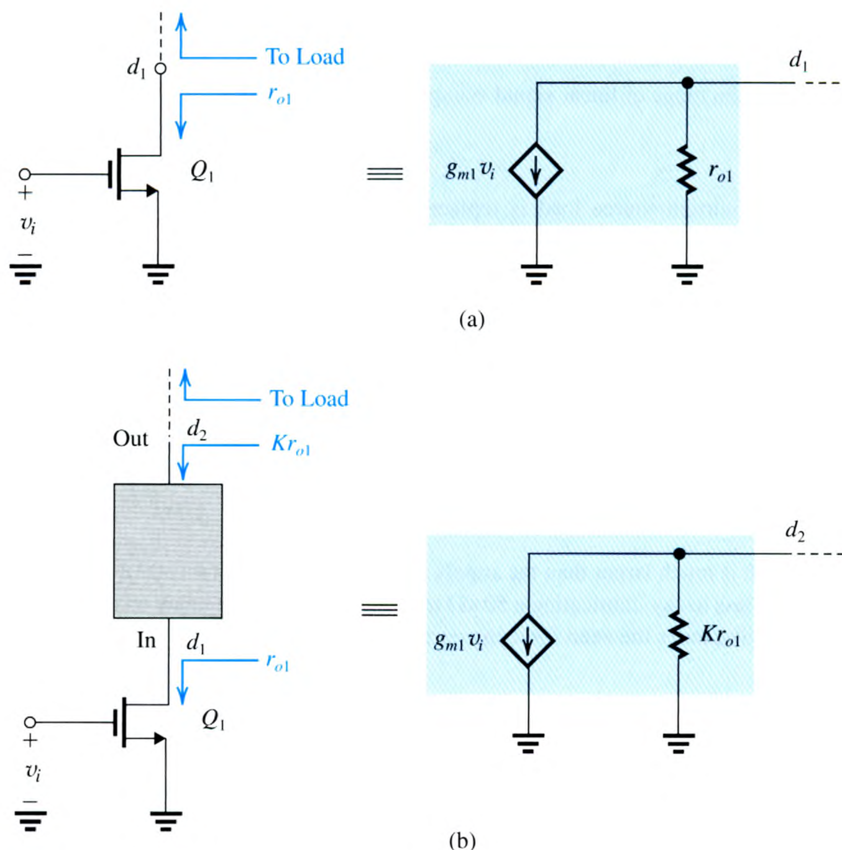


Figure 8.17 To increase the voltage gain realized in the basic gain cell shown in (**a**), a functional block, shown as a black box in (**b**), is connected between d_1 and the load. This new block is required to pass the current $g_{m1}v_i$ right through but raise the resistance level by a factor K. The functional block is a current buffer and can be realized with a common-gate transistor, as demonstrated in the next section.

being we are not showing the load device. In Fig. 8.17(b) we have inserted a shaded box between the drain of Q_1 and a new output terminal labeled d_2. Here again we are not showing the load to which d_2 will be connected. Our "black box" takes in the output current of Q_1 and passes it to the output; thus at its output we have the equivalent circuit shown, consisting of the same controlled source $g_{m1}v_i$ but with the output resistance increased by a factor K.

Now, what does the black box really do? Since it passes the *current* $g_{m1}v_i$ right through, it must have a very low, ideally zero, input resistance, and a current gain of unity. These two properties, together with the fact that it has a high output resistance, mean that it is a **current buffer**. It is the dual of the voltage buffer (the source and emitter followers), which passes the *voltage* but *lowers* the resistance level.

If we search our repertoire of transistor amplifier configurations studied in Section 7.3, the only candidate for implementing this current-buffering action is the common-gate (or in bipolar, common-base) amplifier. Indeed, recall that the CG and CB circuits have a unity current gain. What we have not yet investigated, however, is their resistance transformation property. We shall do this in the next section.

8.4 The Common-Gate and Common-Base Amplifiers as Current Buffers

In this section, we study the IC versions of the CG and CB amplifier configurations focusing on their use as current buffers. This study differs in a significant way from that of the discrete-circuit versions (Section 7.3.5) because here we have to take into account the output resistance of the transistor, r_o.

8.4.1 The CG Circuit

Figure 8.18(a) shows a CG amplifier with the biasing arrangement shown only partially. The amplifier is fed with a current signal source i_{sig} having a resistance R_s, and it has a load resistance R_L. The latter is usually implemented using a PMOS current source, as discussed earlier.

To characterize the signal performance of the CG amplifier as a current buffer, we model it with the equivalent circuit shown in Fig. 8.18(b). Here R_{in} is the input resistance looking into the source of the CG transistor, R_{out} is the output resistance looking back into the drain with i_{sig} set to zero, and A_{is} is the current gain i_o/i_{sig} with R_L set to zero; in other words, A_{is} is the short-circuit current gain of the CG amplifier.

Input Resistance The input resistance R_{in} can be found using the circuit of Fig. 8.19. Here we have employed the π model of the MOSFET and applied a test voltage v_x to the input. The input resistance is given by

$$R_{in} \equiv \frac{v_x}{i_x}$$

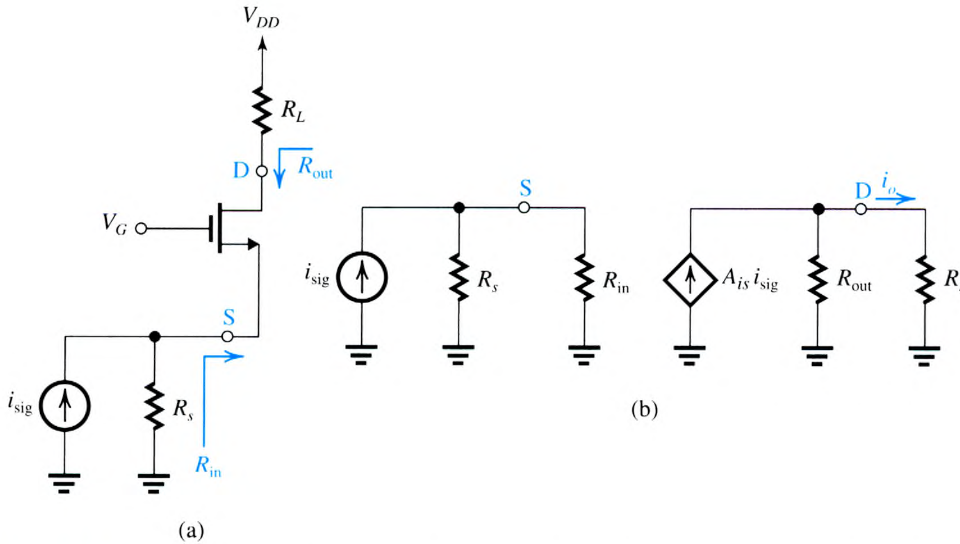


Figure 8.18 (a) A CG amplifier with the bias arrangement only partially shown. (b) Equivalent circuit model of the CG amplifier in (a).

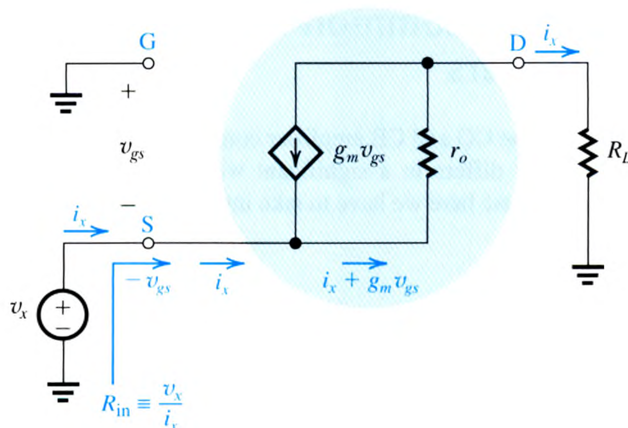


Figure 8.19 Determining the input resistance R_{in} of the CG amplifier.

Some of the analysis is shown in Fig. 8.19. Now, writing a loop equation for the loop comprising v_x, r_o, and R_L gives

$$v_x = (i_x + g_m v_{gs})r_o + i_x R_L$$

Since the voltage at the source node v_x is equal to $-v_{gs}$, we can replace v_{gs} by $-v_x$ and rearrange terms to obtain $R_{in} \equiv v_x/i_x$,

$$R_{in} = \frac{r_o + R_L}{g_m r_o + 1} \quad (8.51)$$

For $g_m r_o \gg 1$,

$$R_{in} \simeq \frac{1}{g_m} + \frac{R_L}{g_m r_o} \quad (8.52)$$

This is a very interesting result. First, it shows that if r_o is infinite, as was the case in our analysis of the discrete CG amplifier in Section 7.3.5, then R_{in} reduces to $1/g_m$, verifying the result we found there. If r_o cannot be neglected, as is always the case in IC amplifiers, we see that the input resistance depends on R_L in an interesting fashion: The load resistance R_L is *transformed* to the input by *dividing* it by the intrinsic gain $A_0 = g_m r_o$. Thus, even as R_L is increased, this impedance transformation property ensures that R_{in} remains relatively low, an important characteristic of a current buffer.

Short-Circuit Current Gain The short-circuit current gain can be determined using the circuit in Fig. 8.20(a) as

$$A_{is} \equiv \frac{i_o}{i_{sig}}$$

Observe that by considering the circuit composed of the controlled source $g_m v_{gs}$ and the resistance r_o as a supernode, the current entering that node will be equal to the current exiting, namely i_o. Next, note that the voltage at the source node S is $-v_{gs}$ and that this voltage appears across the current source $g_m v_{gs}$; thus, we can use the source absorption theorem (Appendix C) to replace the controlled source by a resistance $1/g_m$. This produces the simplified equivalent

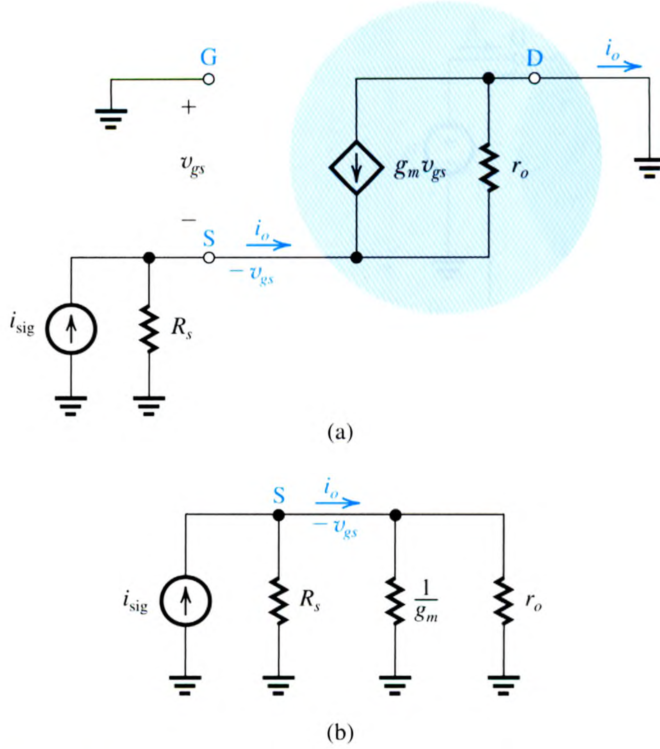


Figure 8.20 **(a)** Determining the short-circuit current gain $A_{is} = i_o/i_{sig}$ of the CG amplifier. **(b)** A simplified version of the circuit in (a).

circuit shown in Fig. 8.20(b), from which we can find i_o/i_{sig} by using the current-divider rule,

$$A_{is} \equiv \frac{i_o}{i_{sig}} = \frac{g_m + \dfrac{1}{r_o}}{g_m + \dfrac{1}{r_o} + \dfrac{1}{R_s}}$$

Thus,

$$A_{is} = \frac{1 + \dfrac{1}{g_m r_o}}{1 + \dfrac{1}{g_m r_o} + \dfrac{1}{g_m R_s}} \tag{8.53}$$

For $g_m r_o \gg 1$ and $g_m R_s \gg 1$,

$$A_{is} \simeq 1$$

which is an important characteristic of a current buffer.

Output Resistance To obtain the output resistance R_{out} we use the circuit shown in Fig. 8.21. Here we have open circuited i_{sig} but left the source resistance R_s, and applied a test voltage v_x to the output. The output resistance is given by

$$R_{out} = \frac{v_x}{i_x}$$

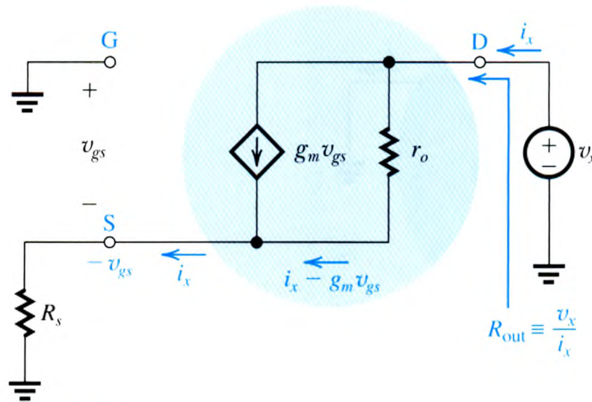


Figure 8.21 Determining the output resistance R_{out} of the CG amplifier.

Some of the analysis is shown in Fig. 8.21. Now, a loop equation for the loop comprising v_x, r_o, and R_s gives

$$v_x = (i_x - g_m v_{gs})r_o + i_x R_s \tag{8.54}$$

Notice that the voltage at the source terminal is $-v_{gs}$ and thus can also be expressed as

$$-v_{gs} = i_x R_s$$

Substituting this value for v_{gs} in Eq. (8.54) and rearranging terms to obtain $R_{out} \equiv v_x/i_x$ yields

$$R_{out} = r_o + R_s + g_m r_o R_s \tag{8.55}$$

which can be written in the alternate form

$$R_{out} = r_o + (1 + g_m r_o)R_s \tag{8.56}$$

For $g_m r_o \gg 1$,

$$R_{out} \simeq r_o + (g_m r_o)R_s \tag{8.57}$$

and if we also have $g_m R_s \gg 1$, then

$$R_{out} \simeq (g_m r_o)R_s \tag{8.58}$$

Equation (8.57) indicates that the output resistance of the CG amplifier includes, in addition to the transistor's r_o, a component related to the resistance in the source lead R_s. The key point is that the CG amplifier *transforms* the source resistance R_s to the output by *multiplying* it by the intrinsic gain $A_0 = g_m r_o$. This impedance transformation is the inverse to that observed from output to input. Now, if R_s is large then the output resistance of the CG circuit can be very large; this also is an important characteristic of a current buffer.

To summarize, the CG circuit has a unity current gain; a low input resistance, obtained by dividing R_L by $g_m r_o$; and a high output resistance, obtained by multiplying R_s by $g_m r_o$. Thus it makes for an excellent current buffer and can be used to implement the shaded functional box in Fig. 8.17. As a useful summary, Fig. 8.22 illustrates the impedance transformation properties of the common-gate amplifier.

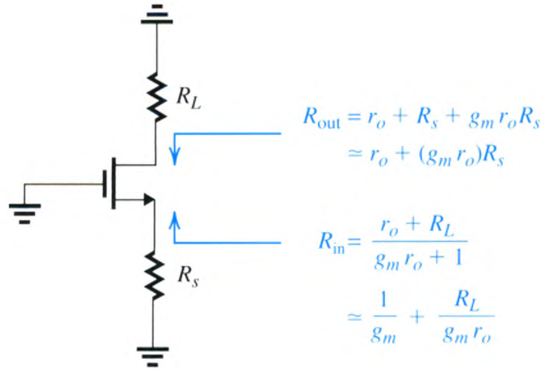


Figure 8.22 The impedance transformation properties of the common-gate amplifier. Depending on the values of R_s and R_L, we can sometimes write $R_{in} \simeq R_L/(g_m r_o)$ and $R_o \simeq (g_m r_o)R_s$. However, such approximations are not always justified.

EXERCISES

8.10 For the CG amplifier in Fig. 8.18(a) with the input signal source replaced by its Thévenin equivalent, show that the voltage gain is given by

$$\frac{v_o}{v_{sig}} = \frac{R_L}{R_s + R_{in}}$$

where v_o is the signal voltage at the drain.

8.11 For a CG amplifier for which $g_m r_o \gg 1$, find R_{in} for the following cases: $R_L = 0$; r_o; $(g_m r_o)r_o$; ∞.
Ans. $\frac{1}{g_m}$; $\frac{2}{g_m}$; r_o; ∞

8.12 For a CG amplifier for which $g_m r_o \gg 1$, find R_{out} for the following cases: $R_s = 0$; r_o; $(g_m r_o)r_o$; ∞.
Ans. r_o; $(g_m r_o)r_o$; $(g_m r_o)^2 r_o$; ∞

8.13 For a CG amplifier for which $R_s = r_o$, find A_{is} for the cases: $g_m r_o = 20$, 50, and 100.
Ans. 0.95 A/A; 0.98 A/A; 0.99 A/A

8.4.2 Output Resistance of a CS Amplifier with a Source Resistance

In Section 7.3.4, we considered some of the benefits of including a resistance R_s in the source lead of a CS amplifier, as in Fig. 8.23. Such a resistance is referred to as a source-degeneration resistance because it reduces the effective transconductance of the CS stage to $g_m/(1 + g_m R_s)$, that is, by a factor $(1 + g_m R_s)$. This also is the factor by which a number of performance parameters are increased, such as linearity and bandwidth (as we will see in Chapter 10). At this point, we simply wish to show that the expression we derived for the output resistance of the CG amplifier applies directly to the case of a source-degenerated CS amplifier. This is because when we determine R_o, we ground the input terminal, making transistor Q appear as a CG transistor. Thus R_o is given by Eq. (8.55), namely,

$$R_o = r_o + R_s + g_m r_o R_s \tag{8.59}$$

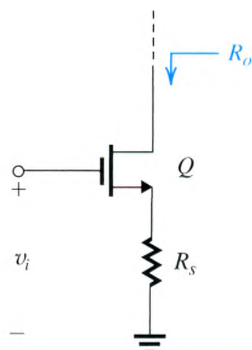


Figure 8.23 The output resistance expression of the CG amplifier can be used to find the output resistance of a source-degenerated common-source amplifier. Here, a useful interpretation of the result is that R_s increases the output resistance by the factor $(1 + g_m R_s)$.

Since manually $g_m r_o \gg 1$, the second term on the right-hand side will be much lower than the third and can be neglected, resulting in

$$R_o \simeq (1 + g_m R_s) r_o \tag{8.60}$$

Thus source degeneration increases the output resistance of the CS amplifier from r_o to $(1 + g_m R_s) r_o$, again by the same factor $(1 + g_m R_s)$. In Chapter 11, we will find that R_s introduces negative (degenerative) feedback of an amount $(1 + g_m R_s)$.

Video Example VE 8.3 MOS Current Source with Source Degeneration

Figure VE8.3 shows a current source realized using a current mirror with two matched transistors Q_1 and Q_2. Two equal resistances R_s are inserted in the source leads to increase the output resistance of the current source. If Q_2 is operating at $g_m = 2$ mA/V and has $V_A = 5$ V, and if the maximum allowed dc voltage drop across R_s is 0.2 V, what is the maximum available output resistance of the current source? Assume that the voltage at the common-gate node is approximately constant.

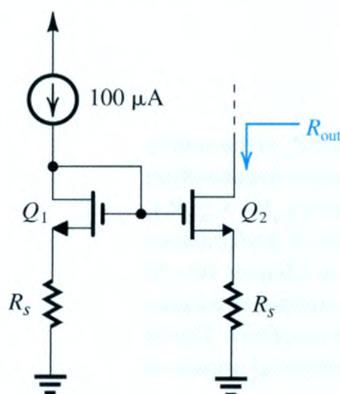


Figure VE8.3 Circuit for Video Example VE8.3.

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.



Related end-of-chapter problem: 8.54

EXERCISE

8.14 Given that source degeneration reduces the transconductance of a CS amplifier from g_m to approximately $g_m/(1+g_m R_s)$ and increases its output resistance by approximately the same factor, what happens to the open-circuit voltage gain A_{vo}? Now, find an expression for A_v when a load resistance R_L is connected to the output.

Ans. A_{vo} remains constant at $g_m r_o$:

$$A_v = (g_m r_o)\frac{R_L}{R_L + (1+g_m R_s)r_o} \qquad \text{(E.8.14)}$$

8.4.3 The Body Effect in the CG Amplifier

Since in the CG amplifier the source cannot be connected to the substrate, the body effect (see Section 5.4) plays a role in the operation of the CG amplifier. It turns out, however, that taking the body effect into account in the analysis of the CG circuit is very simple. To see how it is done, refer to Fig. 8.24(a) and recall that the body terminal acts as another gate for the MOSFET. Thus, just as a signal voltage v_{gs} between the gate and the source gives rise to a drain current signal $g_m v_{gs}$, a signal voltage v_{bs} between the body and the source gives rise to a drain current signal $g_{mb} v_{bs}$. Thus the drain signal current becomes $(g_m v_{gs} + g_{mb} v_{bs})$, where the body transconductance g_{mb} is a small fraction χ of g_m; $g_{mb} = \chi g_m$ and $\chi = 0.1$ to 0.2. For the CG circuit, $v_{bs} = v_{gs}$, the two current signals can thus be combined as $(g_m + g_{mb})v_{gs}$ or $g_m(1+\chi)v_{gs}$. The body effect can then be taken into account by simply replacing g_m by $g_m(1+\chi)$ as shown in the π equivalent model shown in Fig. 8.24(b). Normally, however, we will not bother with the factor $(1+\chi)$ in our calculations.

8.4.4 The CB Circuit

Analysis of the CB amplifier parallels that of the CG amplifier except that it is complicated by the finite base current, which manifests itself as a resistance r_π between the base and emitter terminals.

Figure 8.25(a) shows a CB amplifier with the bias details only partially shown and with a load resistance R_L that is normally implemented with a *pnp* current source. To characterize the

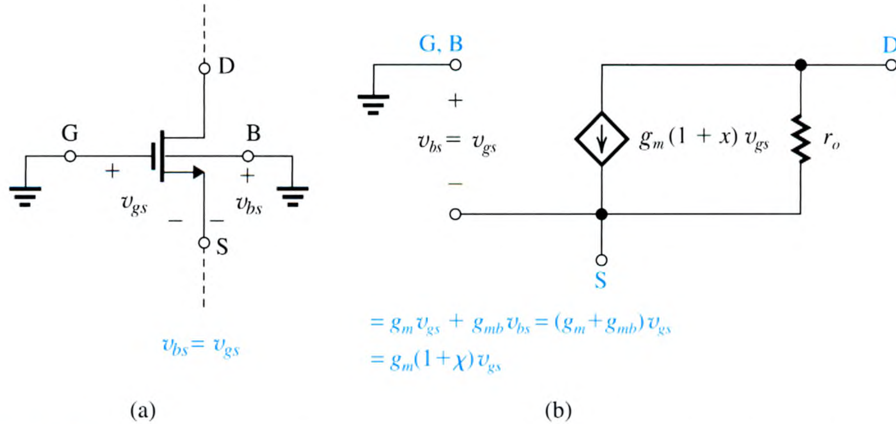


Figure 8.24 The body effect can be easily taken into account in the analysis of the CG circuit by replacing g_m by $(1+\chi)g_m$, where $\chi = g_{mb}/g_m = 0.1$ to 0.2.

signal performance of the CB amplifier as a current buffer, we model it with the equivalent circuit shown in Fig. 8.25(b), which is the same model used for the CG circuit above. To determine the model parameters, R_{in}, A_{is}, and R_{out}, we replace the BJT with its hybrid-π model and obtain the circuit shown in Fig. 8.25(c). Here, we observe that unlike the CG case, we have an additional resistance r_π that appears between emitter and base and thus in parallel with R_e. Using this observation, we can obtain expressions for R_{in}, A_{is}, and R_{out} by appropriately modifying the corresponding expressions derived above for the CG case.

We can find the input resistance R_{in} of the CB amplifier by augmenting the expression for R_{in} of the CG circuit in Eq. (8.51) with a parallel resistance r_π, thus

$$R_{in} = r_\pi \,\|\, [\frac{r_o + R_L}{g_m r_o + 1}] \tag{8.61}$$

For $g_m r_o \gg 1$,

$$R_{in} \simeq r_\pi \,\|\, [\frac{1}{g_m} + \frac{R_L}{g_m r_o}] \tag{8.62}$$

Here we see that $r_o = \infty$ yields $R_{in} = r_\pi \,\|\, \frac{1}{g_m} = r_e$, which is consistent with the case of the discrete-circuit CB amplifier studied in Section 7.3.5. Also, for $R_L = 0$, $R_{in} = r_e$. As R_L is increased from 0 and as long as r_π is much greater than the value between the square brackets,

$$R_{in} \simeq \frac{1}{g_m} + \frac{R_L}{g_m r_o} \tag{8.63}$$

which is identical to the CG case, indicating that the CB amplifier has similar resistance-transformation properties. However, unlike the CG amplifier, the CB circuit has a maximum value for R_{in} determined by the shunt resistance r_π,

$$R_{inmax} = r_\pi \tag{8.64}$$

which is reached when $R_L = \infty$.

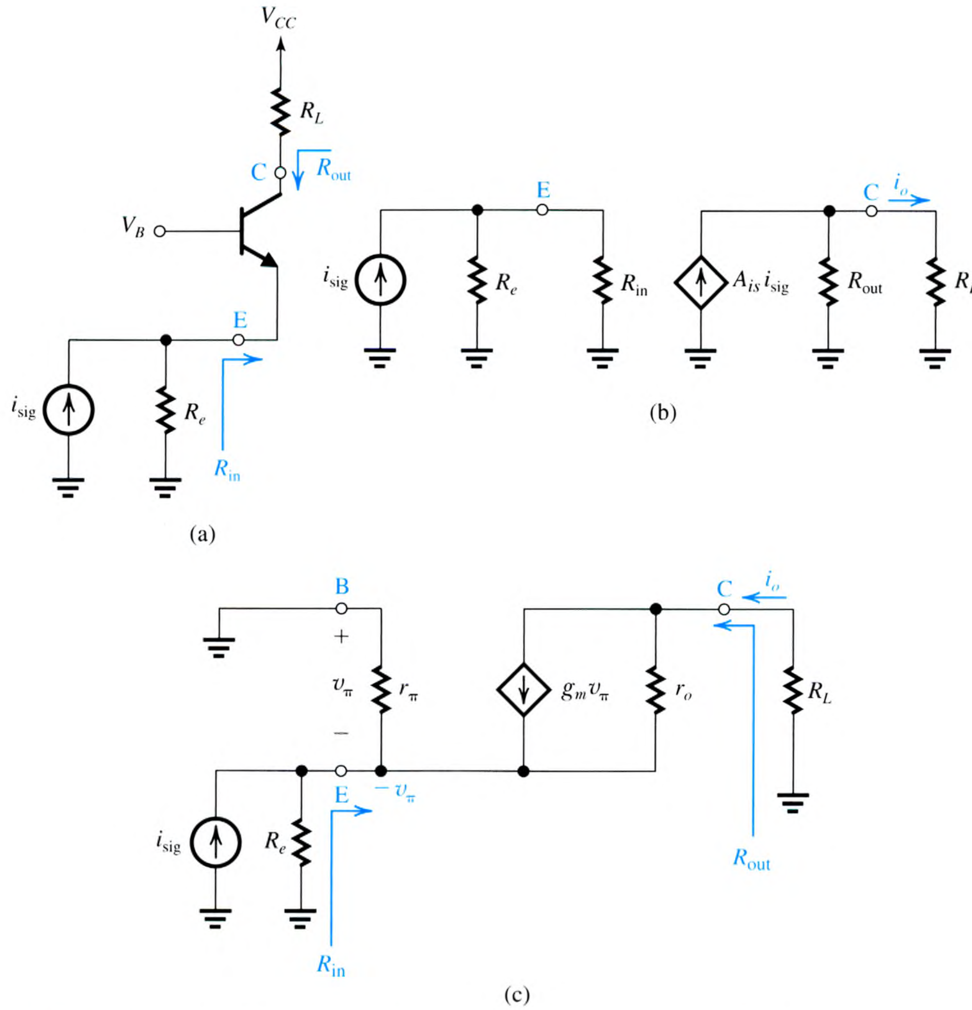


Figure 8.25 (**a**) A CB amplifier. (**b**) Equivalent circuit model of the CB amplifier in (a). (**c**) The CB amplifier in (a) with the transistor replaced with its π model.

We can find the short-circuit current gain of the CB amplifier by replacing R_s in Eq. (8.53) with $(R_e \| r_\pi)$, thus

$$A_{is} = \frac{1 + \dfrac{1}{g_m r_o}}{1 + \dfrac{1}{g_m r_o} + \dfrac{1}{g_m (R_e \| r_\pi)}} \tag{8.65}$$

which for $g_m r_o \gg 1$ and $R_e \gg r_\pi$ reduces to

$$A_{is} \simeq \frac{1}{1 + \dfrac{1}{g_m r_\pi}} = \frac{1}{1 + \dfrac{1}{\beta}} = \alpha \tag{8.66}$$

For $\beta \gg 1$,

$$A_{is} \simeq 1$$

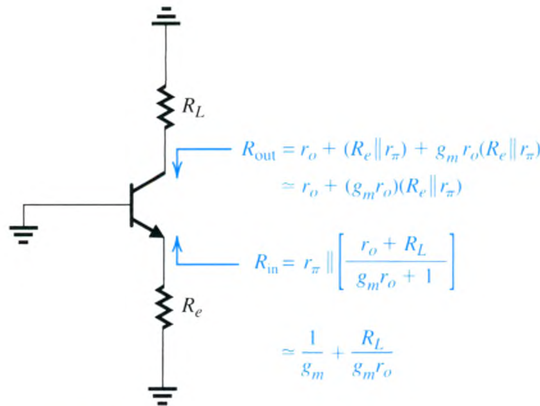


Figure 8.26 The impedance transformation properties of the CB amplifier. Note that for $\beta = \infty$, these formulas reduce to those for the MOSFET case (Fig. 8.21).

The output resistance R_{out} of the CB amplifier can be obtained by replacing R_s in Eq. (8.55) with $(R_e \| r_\pi)$, thus

$$R_{out} = r_o + (R_e \| r_\pi) + g_m r_o (R_e \| r_\pi) \tag{8.67}$$

which can be written in the alternative form

$$R_{out} = r_o + (1 + g_m r_o)(R_e \| r_\pi) \tag{8.68}$$

For $g_m r_o \gg 1$,

$$R_{out} \simeq r_o + (g_m r_o)(R_e \| r_\pi) \tag{8.69}$$

Thus, similar to the CG amplifier, the CB amplifier exhibits an impedance transformation property that raises the output resistance. Unlike the CG case, however, the output resistance of the CB circuit has an absolute maximum value obtained by setting $R_e = \infty$ as

$$R_{out}\big|_{max} = r_o + g_m r_o r_\pi = (\beta + 1) r_o \tag{8.70}$$

We conclude that the CB circuit has a current gain of nearly unity, a low input resistance, and a high output resistance; thus it makes for an excellent current buffer. The impedance transformation properties of the CB circuit are summarized in Fig. 8.26.

EXERCISES

8.15 For a CB amplifier with $g_m r_o \gg 1$, find approximate values for R_{in} for the following cases: $R_L = 0$; r_o; βr_o; ∞.
Ans. r_e; $2r_e$; $\frac{1}{2} r_\pi$; r_π

8.16 For a CB amplifier with $g_m r_o \gg 1$, find approximate values for R_{out} for the following cases: $R_e = 0$; r_e; r_π; r_o; ∞.
Ans. r_o; $2r_o$; $(\frac{1}{2}\beta + 1)r_o$; $(\beta + 1)r_o$; $(\beta + 1)r_o$

8.17 For a CB amplifier with $R_e = r_o$ and $\beta = 100$, find A_{is} for $g_m r_o = 100$ and 1000.
Ans. 0.98 A/A; 0.99 A/A

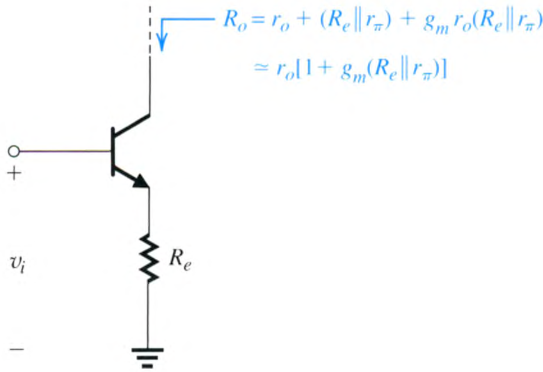


Figure 8.27 Output resistance of a CE amplifier with an emitter resistance R_e.

8.4.5 Output Resistance of the Emitter-Degenerated CE Amplifier

As in the MOS case, we can use the expression for the output resistance derived for the CB amplifier (Eq. 8.69) for the case of a CE amplifier with a resistance R_e connected in its emitter, as shown in Fig. 8.27,

$$R_o \simeq r_o + g_m r_o (R_e \| r_\pi) \tag{8.71}$$

which can be written in the alternate form

$$R_o = [1 + g_m (R_e \| r_\pi)] r_o \tag{8.72}$$

Thus, emitter degeneration multiplies the transistor output resistance r_o by the factor $[1 + g_m (R_e \| r_\pi)]$. Note that this factor has a maximum value of $(1 + g_m r_\pi)$ or $(\beta + 1)$, obtained when $R_e \gg r_\pi$. Thus the theoretical maximum output resistance realized is $(\beta + 1) r_o$ and is achieved when the emitter is open circuited.

EXERCISE

8.18 Find the output resistance of a CE amplifier biased at $I_C = 0.1$ mA and having a resistance of 1 $k\Omega$ connected in its emitter. Let $\beta = 100$ and $V_A = 20$ V. What is the value of the output resistance without degeneration?
Ans. 1 MΩ; 200 kΩ

8.5 The Cascode Amplifier

Cascoding refers to the use of a transistor connected in the common-gate (or the common-base) configuration to provide **current buffering** for the output of a common-source (or a common-emitter) amplifying transistor. As we found from our study of the CG and CB circuits in the previous section, their current-buffering action provides a high output resistance, thus enabling the cascode amplifier to have a voltage gain much larger than that possible with the CS or CE gain cells alone.

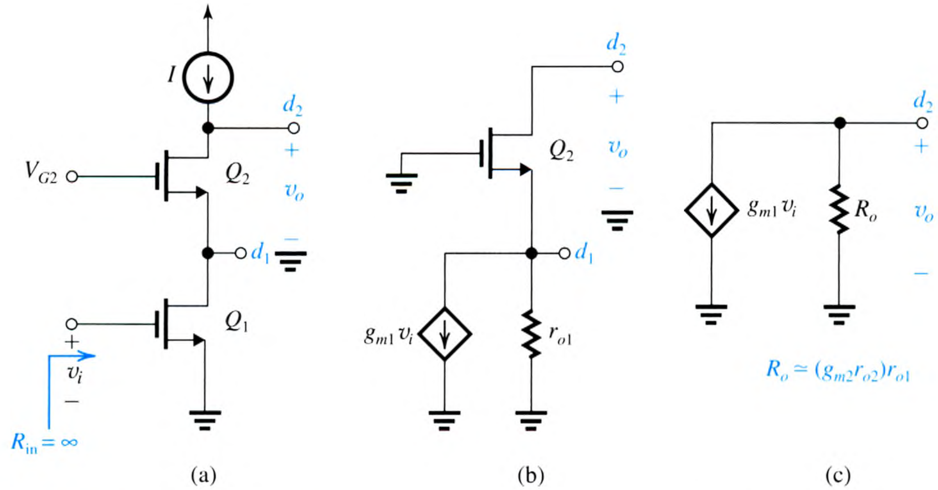


Figure 8.28 **(a)** A cascode amplifier with a constant-current source load. **(b)** The amplifier equivalent circuit with Q_1 replaced with its output equivalent circuit, and the dc sources eliminated. Note that R_L of the cascode amplifier is infinite. **(c)** The output equivalent circuit of the cascode amplifier.

8.5.1 The MOS Cascode Amplifier

The Ideal Case Figure 8.28(a) shows a MOS cascode amplifier loaded with an ideal constant-current source I. Here Q_1 is the CS amplifying transistor, and Q_2 is the CG cascode[2] transistor. Since the constant-current source is ideal, its incremental resistance will be infinite, and thus the cascode amplifier will be operating with an infinite load resistance. It follows that the voltage gain realized, v_o/v_i, will be the open-circuit voltage gain

$$A_{vo} \equiv \frac{v_o}{v_i}$$

To find A_{vo}, we first replace Q_1 with its output equivalent circuit consisting of the controlled current source $g_{m1}v_i$ and the output resistance r_{o1}, as shown in Fig. 8.28(b). Next, recalling that the CG transistor has a short-circuit current gain of almost unity and an output resistance that is approximately $(g_{m2}r_{o2})$ times the resistance in its source lead, r_{o1}, we can say that the output equivalent circuit will be as shown in Fig. 8.28(c).

$$R_o \simeq (g_{m2}r_{o2})r_{o1} \tag{8.73}$$

The open-circuit voltage gain can now be determined as follows,

$$\begin{aligned} v_o &= -g_{m1}v_iR_o \\ A_{vo} &\equiv \frac{v_o}{v_i} = -g_{m1}R_o \\ &= -(g_{m1}r_{o1})(g_{m2}r_{o2}) \end{aligned} \tag{8.74}$$

[2]The name *cascode* is a carryover from the days of vacuum tubes and is a shortened version of "*casc*aded cath*ode*"; in the tube version, the anode of the amplifying tube (corresponding to the drain of Q_1) feeds the cathode of the cascode tube (corresponding to the source of Q_2).

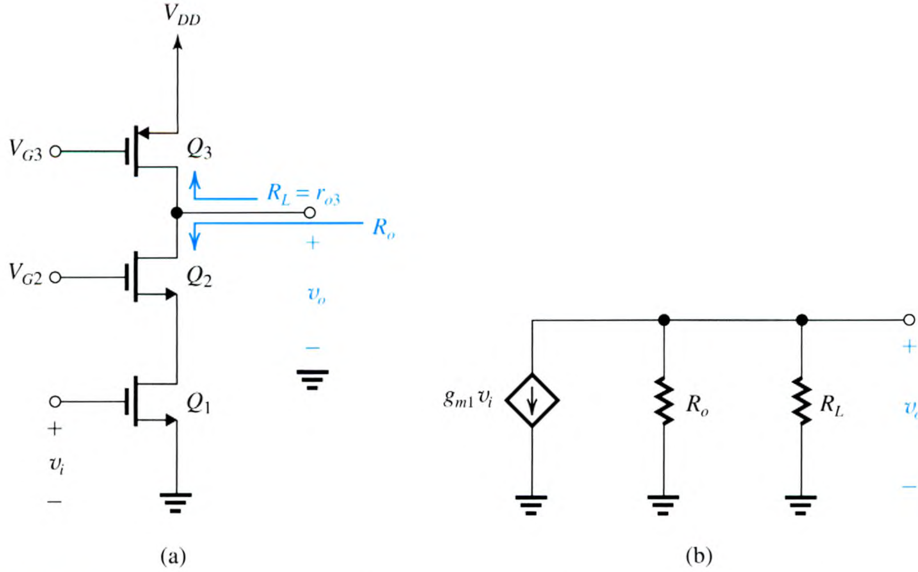


Figure 8.29 (a) A MOS cascode amplifier loaded in a simple PMOS current source Q_3. (b) Equivalent circuit at the amplifier output.

For the case $g_{m1} = g_{m2} = g_m$ and $r_{o1} = r_{o2} = r_o$,

$$A_{vo} = -(g_m r_o)^2 = -A_0^2 \tag{8.75}$$

Thus cascoding increases the gain magnitude from A_0 to A_0^2.

Implementation of the Constant-Current Source Load If the current source load is implemented with a PMOS transistor (which can be part of a PMOS current mirror) as shown in Fig. 8.29(a), the load resistance R_L will be equal to the output resistance of Q_3, r_{o3},

$$R_L = r_{o3}$$

and the voltage gain of the cascode amplifier will be

$$\begin{aligned} A_v &= -g_{m1}(R_o \| R_L) \\ &= -g_{m1}(g_{m2}r_{o2}r_{o1} \| r_{o3}) \end{aligned} \tag{8.76}$$

from which we can readily see that since $R_L \ll R_o$, the total resistance will be approximately equal to r_{o3} and the gain will be

$$A_v \simeq -g_{m1}r_{o3} \tag{8.77}$$

Thus the gain magnitude will be back to A_0, of the same order as that realized by a CS amplifier. In other words, the use of a simple current-source load with a relatively low output resistance has in effect destroyed the cascoding advantage of increased output resistance. Nevertheless, it turns out that this cascode amplifier, whose gain is of the same order as that of a CS amplifier, does in fact have a major advantage over the CS circuit: It exhibits a much wider bandwidth. We will demonstrate this point in Chapter 10.

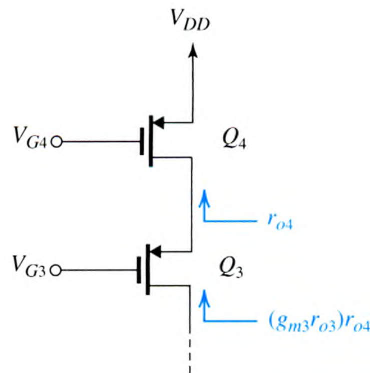
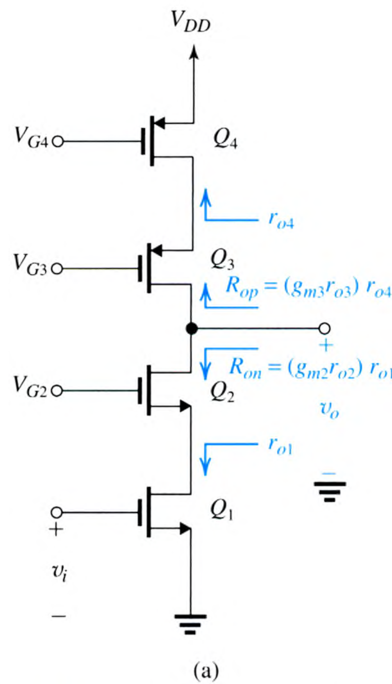


Figure 8.30 Employing a cascode transistor Q_3 to raise the output resistance of the current source Q_4.



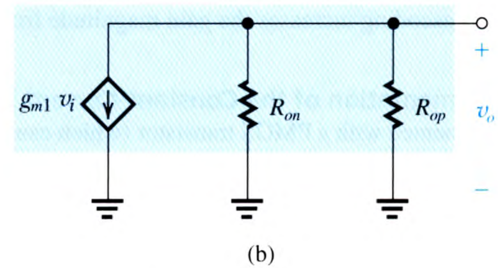


Figure 8.31 A cascode amplifier with a cascode current-source load.

The Use of a Cascode Current Source To realize a gain of the order of A_0^2, the load resistance R_L must be of the same order as R_o of the cascode amplifier. We can do this using a cascode current source like the one in Fig. 8.30. Here Q_4 is the current-source transistor, and Q_3 is the CG cascode transistor. Voltages V_{G3} and V_{G4} are dc bias voltages. The cascode transistor Q_3 multiplies the output resistance of Q_4, r_{o4}, by $(g_{m3}r_{o3})$ to provide an output resistance for the cascode current source of

$$R_o = (g_{m3}r_{o3})r_{o4} \tag{8.78}$$

Combining a cascode amplifier with a cascode current source results in the circuit of Fig. 8.31(a). The equivalent circuit at the output side is shown in Fig. 8.31(b), from which

we can easily find the voltage gain:

$$A_v = \frac{v_o}{v_i} = -g_{m1}\left[R_{on} \parallel R_{op}\right]$$

Thus,

$$A_v = -g_{m1}\left\{\left[(g_{m2}r_{o2})r_{o1}\right] \parallel \left[(g_{m3}r_{o3})r_{o4}\right]\right\} \tag{8.79}$$

For the case in which all transistors are identical,

$$A_v = -\frac{1}{2}(g_m r_o)^2 = -\frac{1}{2}A_0^2 \tag{8.80}$$

By comparison to the gain expression in Eq. (8.50), we see that using the cascode configuration for both the amplifying transistor and the current-source load transistor results in an increase in the magnitude of gain by a factor equal to A_0.

Example 8.4

Design the cascode current source of Fig. 8.30 to provide a current of 100 μA and an output resistance of 500 kΩ. Assume the availability of a 0.18-μm CMOS technology for which $V_{DD} = 1.8$ V, $V_{tp} = -0.5$ V, $\mu_p C_{ox} = 90\ \mu\text{A/V}^2$, and $V_A' = -5$ V/μm. Use $|V_{OV}| = 0.3$ V and determine L and W/L for each transistor, and the values of the bias voltages V_{G3} and V_{G4}. Design to maximize the allowable signal swing at the output terminal.

Solution

The output resistance R_o is given by

$$R_o = (g_{m3}r_{o3})r_{o4}$$

Assuming Q_3 and Q_4 are identical,

$$R_o = (g_m r_o)r_o = \frac{|V_A|}{|V_{OV}|/2} \times \frac{|V_A|}{I_D}$$

Using $|V_{OV}| = 0.3$ V, we write

$$500\ \text{k}\Omega = \frac{|V_A|}{0.15} \times \frac{|V_A|}{0.1\ \text{mA}}$$

Thus we require

$$|V_A| = 2.74\ \text{V}$$

Now, since $|V_A| = |V_A'|\,L$ we need to use a channel length of

$$L = \frac{2.74}{5} = 0.55\ \mu\text{m}$$

Example 8.4 *continued*

which is about three times the minimum channel length. With $|V_t| = 0.5$ V and $|V_{OV}| = 0.3$ V,

$$V_{SG4} = 0.5 + 0.3 = 0.8\text{ V}$$

and thus,

$$V_{G4} = 1.8 - 0.8 = 1.0\text{ V}$$

To allow for the largest possible signal swing at the output terminal, we shall use the minimum required voltage across Q_4, namely, $|V_{OV}|$ or 0.3 V. Thus,

$$V_{D4} = 1.8 - 0.3 = 1.5\text{ V}$$

Since the two transistors are identical and are carrying equal currents,

$$V_{SG3} = V_{SG4} = 0.8\text{ V}$$

Thus,

$$V_{G3} = 1.5 - 0.8 = +0.7\text{ V}$$

We note that the maximum voltage allowed at the output terminal of the current source will be constrained by the need to allow a minimum voltage of $|V_{OV}|$ across Q_3; thus;

$$v_{D3\,\max} = 1.5 - 0.3 = +1.2\text{ V}$$

To determine the required W/L ratios of Q_3 and Q_4, we use

$$I_D = \frac{1}{2}(\mu_p C_{ox})\left(\frac{W}{L}\right)|V_{OV}|^2\left(1 + \frac{V_{SD}}{|V_A|}\right)$$

$$100 = \frac{1}{2} \times 90 \times \left(\frac{W}{L}\right) \times 0.3^2\left(1 + \frac{0.3}{2.74}\right)$$

which yields

$$\frac{W}{L} = 22.3$$

Video Example VE 8.4 Design of a MOS Cascode Amplifier

Design the cascode amplifier of Fig. 8.28(a) to obtain $g_{m1} = 2$ mA/V and $R_o = 200$ kΩ. Use the 0.18-μm technology whose parameters are specified in Table K.1 (Appendix K). Determine L, W/L, V_{G2}, and I. Use identical transistors operated at $V_{OV} = 0.2$ V, and design for the maximum possible negative signal swing at the output. What is the value of the minimum permitted output voltage?



VE 8.4

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 8.66

EXERCISES

D8.19 If in Example 8.4, L of each of Q_3 and Q_4 is halved while W/L is changed to allow I_D and V_{OV} to remain unchanged, find the new values of R_o and W/L. [*Hint*: In computing the required (W/L), note that $|V_A|$ has changed.]
Ans. 126 kΩ; 20.3

8.20 Consider the cascode amplifier of Fig. 8.31 with the dc component at the input, $V_I = 0.7$ V, $V_{G2} = 1.0$ V, $V_{G3} = 0.8$ V, $V_{G4} = 1.1$ V, and $V_{DD} = 1.8$ V. If all devices are matched (i.e., if $k_{n1} = k_{n2} = k_{p3} = k_{p4}$), and have equal $|V_t|$ of 0.5 V, what is the overdrive voltage at which the four transistors are operating? What is the allowable voltage range at the output?
Ans. 0.2 V; 0.5 V to 1.3 V

8.21 The cascode amplifier in Fig. 8.31 is operated at a current of 0.2 mA with all devices operating at $|V_{OV}| = 0.2$ V. All devices have $|V_A| = 2$ V. Find g_{m1}, the output resistance of the amplifier, R_{on}, and the output resistance of the current source, R_{op}. Also find the overall output resistance and the voltage gain realized.
Ans. 2 mA/V; 200 kΩ, 200 kΩ; 100 kΩ; −200 V/V

8.5.2 Distribution of Voltage Gain in a Cascode Amplifier

It is often useful to know how much of the overall voltage gain of a cascode amplifier is realized in each of its two stages: the CS stage Q_1, and the CG stage Q_2. For this purpose, consider the cascode amplifier shown in Fig. 8.32(a). Here, we have included a load resistance R_L, which represents the output resistance of the current-source load plus any additional resistance that may be connected to the output node. The voltage gain A_v of the amplifier can be found as

$$A_v = -g_{m1}(R_o \parallel R_L)$$

Thus,

$$A_v = -g_{m1}(g_{m2}r_{o2}r_{o1} \parallel R_L) \tag{8.81}$$

The overall gain A_v can be expressed as the product of the voltage gains of Q_1 and Q_2 as

$$A_v = A_{v1}A_{v2} = \left(\frac{v_{o1}}{v_i}\right)\left(\frac{v_o}{v_{o1}}\right) \tag{8.82}$$

To obtain $A_{v1} \equiv v_{o1}/v_i$, we need to find the total resistance between the drain of Q_1 and ground. Referring to Fig. 8.32(b) and denoting this resistance R_{d1}, we can express A_{v1} as

$$A_{v1} = \frac{v_{o1}}{v_i} = -g_{m1}R_{d1} = -g_{m1}(r_{o1} \parallel R_{\text{in}2}) \tag{8.83}$$

where $R_{\text{in}2}$ is the input resistance of the CG transistor Q_2. From Eq. (8.52), we can write

$$R_{\text{in}2} \simeq \frac{R_L}{g_{m2}r_{o2}} + \frac{1}{g_{m2}} \tag{8.84}$$

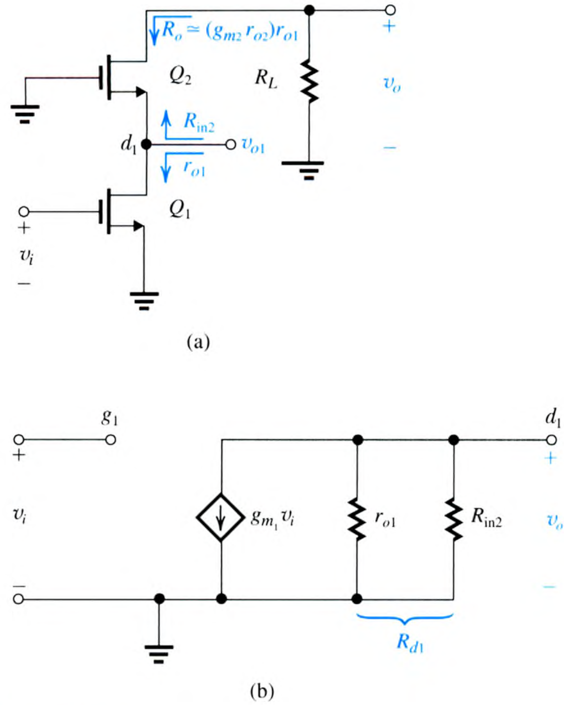


Figure 8.32 (**a**) The cascode amplifier with a load resistance R_L. Only signal quantities are shown. (**b**) Determining v_{o1}.

Table 8.1 Gain Distribution in the MOS Cascode Amplifier for Various Values of R_L

Case	R_L	R_{in2}	R_{d1}	A_{v1}	A_{v2}	A_v
1	∞	∞	r_o	$-g_m r_o$	$g_m r_o$	$-(g_m r_o)^2$
2	$(g_m r_o) r_o$	r_o	$r_o/2$	$-\frac{1}{2}(g_m r_o)$	$g_m r_o$	$-\frac{1}{2}(g_m r_o)^2$
3	r_o	$\frac{2}{g_m}$	$\frac{2}{g_m}$	-2	$\frac{1}{2}(g_m r_o)$	$-(g_m r_o)$
4	$0.2r_0$	$\frac{1.2}{g_m}$	$\frac{1.2}{g_m}$	-1.2	$0.17(g_m r_o)$	$-0.2(g_m r_o)$

Finally, we can obtain A_{v2} by dividing the total gain A_v given by Eq. (8.81) by A_{v1}. For insight into the effect of the value of R_L on the overall gain of the cascode and on how this gain is distributed among the two stages of the cascode amplifier, Table 8.1 provides approximate values for the case $r_{o1} = r_{o2} = r_o$ and for four different values of R_L: (1) $R_L = \infty$, obtained with an ideal current-source load; (2) $R_L = (g_m r_o) r_o$, obtained with a cascode current-source load; (3) $R_L = r_o$, obtained with a simple current-source load; and (4) $R_L = 0.2r_0$, a representative value for some RF amplifier applications.

Observe that while case 1 represents an idealized situation, it is useful in providing the theoretical maximum voltage gain we can achieve in a MOS cascode amplifier. Case 2, which assumes a cascode current-source load with an output resistance equal to that of the cascode amplifier, provides a realistic estimate of the gain we could achieve if we wanted to maximize the realized gain. In certain situations, however, that is not our objective. This point is important, for as we shall see in Section 10.4, there is an entirely different application of the cascode amplifier: to obtain wideband amplification by extending the upper-3-dB frequency f_H. As we will see, this application calls for the situation represented by case 3, where the gain achieved in the CS amplifier is only -2 V/V, and of course the overall gain is now only $-(g_m r_o)$. In Section 10.4 we will see that this trade-off of the overall gain to obtain extended bandwidth is in some cases a good bargain!

EXERCISE

8.22 Consider a cascode amplifier for which the CS and CG transistors are identical and are biased to operate at $I_D = 0.1$ mA with $V_{OV} = 0.2$ V. Also let $V_A = 2$ V. Find A_{v1}, A_{v2}, and A_v for two cases: (a) $R_L = 20$ kΩ and (b) $R_L = 400$ kΩ.
Ans. (a) –1.82 V/V, 10.5 V/V, –19.0 V/V; (b) –10.2 V/V, 19.6 V/V, –200 V/V

8.5.3 The BJT Cascode

Figure 8.33(a) shows the BJT cascode amplifier with an ideal current-source load. Voltage V_{B2} is a dc bias voltage for the CB cascode transistor Q_2. The circuit is very similar to the MOS cascode, and the small-signal analysis will follow in a parallel fashion. First, note that the input resistance of the bipolar cascode amplifier is finite,

$$R_{\text{in}} = r_{\pi 1} \tag{8.85}$$

The equivalent circuit of the output of the cascode amplifier is shown in Fig. 8.33(b). To obtain R_o we use the formula in Eq. (8.68) and note that the resistance R_e in the emitter of Q_2 is r_{o1}, thus

$$R_o \simeq (g_{m2} r_{o2})(r_{o1} \| r_{\pi 2}) \tag{8.86}$$

This result is similar but certainly *not identical* to that for the MOS cascode. Here, because of the finite β of the BJT, we have $r_{\pi 2}$ appearing in parallel with r_{o1}. This poses a very significant constraint on R_o of the BJT cascode. Specifically, because $(r_{o1} \| r_{\pi 2})$ will always be lower than $r_{\pi 2}$, it follows that the maximum possible value of R_o is

$$\begin{aligned} R_o\big|_{\max} &= g_{m2} r_{o2} r_{\pi 2} \\ &= (g_{m2} r_{\pi 2}) r_{o2} = \beta_2 r_{o2} \end{aligned} \tag{8.87}$$

Thus the maximum output resistance we can achieve by cascoding is $\beta_2 r_{o2}$.

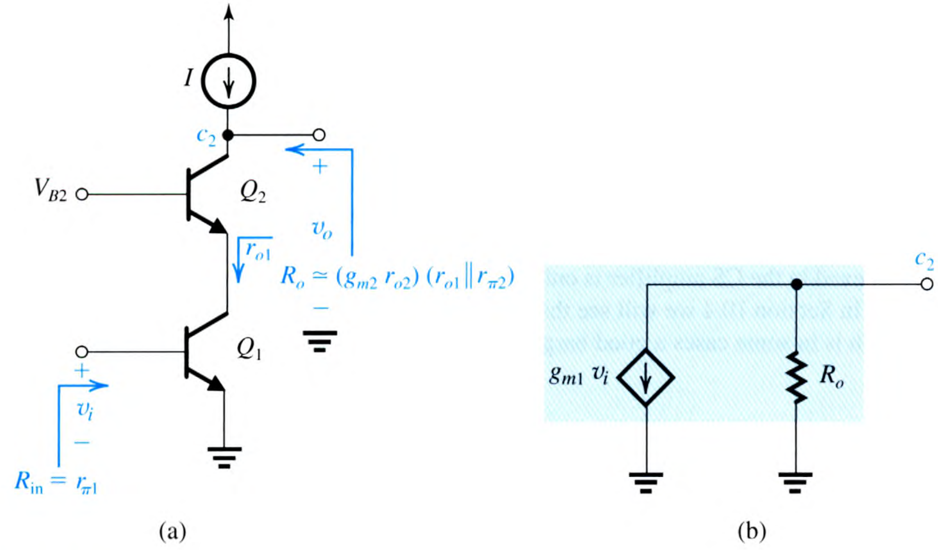


Figure 8.33 (**a**) A BJT cascode amplifier with an ideal current-source load; (**b**) small-signal equivalent-circuit representation of the output of the cascode amplifier.

The open-circuit voltage gain of the bipolar cascode can be found using the equivalent circuit of Fig. 8.33(b) as

$$A_{vo} = \frac{v_o}{v_i} = -g_{m1} R_o$$

Thus,

$$A_{vo} = -g_{m1}(g_{m2} r_{o2})(r_{o1} \| r_{\pi 2}) \tag{8.88}$$

For the case $g_{m1} = g_{m2}$, $r_{o1} = r_{o2}$,

$$A_{vo} = -(g_m r_o)[g_m(r_o \| r_\pi)] \tag{8.89}$$

which will be less than $(g_m r_o)^2$ in magnitude. In fact, the maximum possible gain magnitude is obtained when $r_o \gg r_\pi$ and is given by

$$|A_{vo}|_{\max} = \beta g_m r_o = \beta A_0 \tag{8.90}$$

Note that to realize gains approaching this level, the current-source load must also be cascoded. Figure 8.34 shows a cascode BJT amplifier with a cascode current-source load.

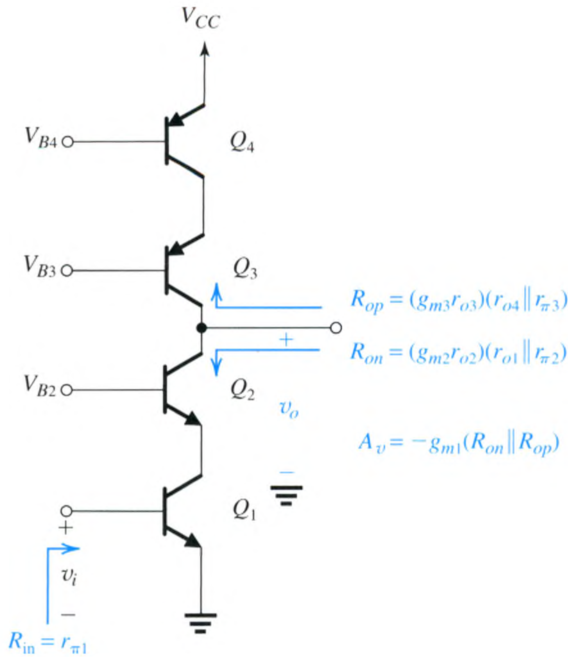


Figure 8.34 A BJT cascode amplifier with a cascode current source.

EXERCISES

8.23 Find an expression for the maximum voltage gain achieved in the amplifier of Fig. 8.34.
Ans. $|A_{v\max}| = g_{m1}(\beta_2 r_{o2} \| \beta_3 r_{o3})$

8.24 Consider the BJT cascode amplifier of Fig. 8.34 when biased at a current of 0.2 mA. Assuming that *npn* transistors have $\beta = 100$ and $V_A = 5$ V and that *pnp* transistors have $\beta = 50$ and $|V_A| = 4$ V, find R_{on}, R_{op}, and A_v. Use the result of Exercise 8.23 to determine the maximum achievable gain.
Ans. 1.67 MΩ; 0.762 MΩ; −4186 V/V; −5714 V/V

8.6 The IC Source Follower

In this section, we study another important building block of integrated-circuit amplifiers: the IC source follower. Having an infinite input resistance and a low-output resistance, the source follower is used as a voltage buffer. Recall that we studied the discrete-circuit source follower in Section 7.3.6.

Figure 8.35(a) shows a source follower formed by transistor Q_1 and biased by a constant-current supplied by the current mirror Q_2–Q_3. Observe that since the source of Q_1 cannot be connected to the body (which is at signal ground potential) a voltage signal v_{bs}

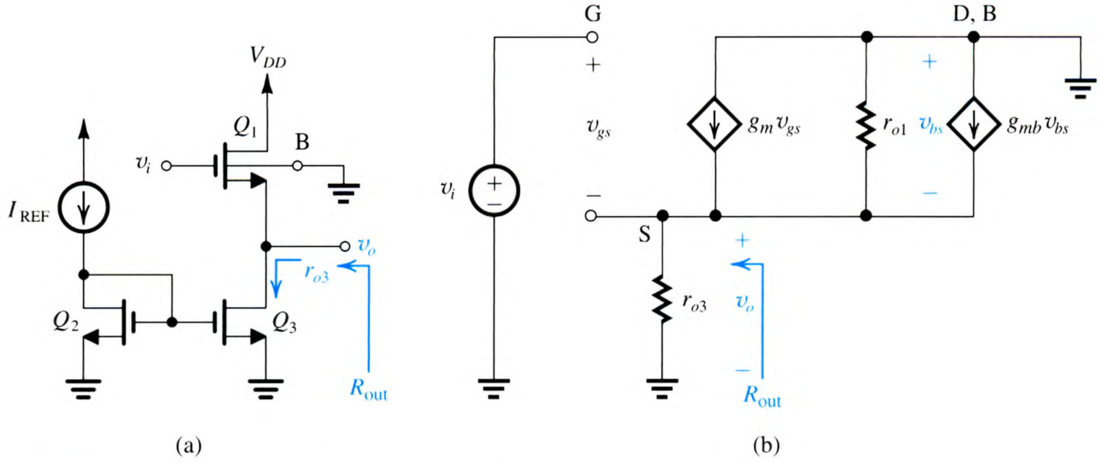


Figure 8.35 (**a**) A source follower biased with a current mirror Q_2–Q_3 and with the body terminal shown. Because the source cannot be connected to the body, the body effect should be taken into account. (**b**) Equivalent circuit.

develops between body and source and gives rise to a current signal $g_{mb}v_{bs}$, as indicated in the equivalent circuit in Fig. 8.35(b). The equivalent circuit shows also the output resistance r_{o3} of the bias current source Q_3, which acts as a load resistance for the follower Q_1.

An important observation to make from the equivalent circuit is that the controlled source ($g_{mb}v_{bs}$) appears across its control voltage v_{bs}. Thus we can use the source-absorption theorem (Appendix D) to replace the controlled source with a resistance $1/g_{mb}$. Next, note that the three resistances $1/g_{mb}$, r_{o1}, and r_{o3} appear in parallel between the source and ground. If we denote their parallel equivalent R_L, we can show that the voltage gain of the source follower is

$$\frac{v_o}{v_i} = \frac{R_L}{R_L + \dfrac{1}{g_m}} \tag{8.91}$$

where

$$R_L = r_{o1} \,\|\, r_{o3} \,\|\, \frac{1}{g_{mb}} \tag{8.92}$$

In cases where $\dfrac{1}{g_{mb}} \ll r_{o1}, r_{o3}$,

$$R_L \simeq \frac{1}{g_{mb}}$$

and

$$\frac{v_o}{v_i} \simeq \frac{g_{mb}}{g_m + g_{mb}} \tag{8.93}$$

Substituting for $g_{mb} = \chi g_m$ where $\chi = 0.1$ to 0.2,

$$\frac{v_o}{v_i} \simeq \frac{1}{1+\chi} \tag{8.94}$$

This is the maximum possible gain obtained from an IC source follower. The actual gain will usually be lower because of the effect of r_{o1} and r_{o3}.

The output resistance R_{out} of the source follower can be obtained from Fig. 8.35(b) as

$$R_{\text{out}} = \frac{1}{g_m} \parallel \frac{1}{g_{mb}} \parallel r_{o1} \parallel r_{o3}$$

$$\simeq \frac{1}{(g_m + g_{mb})} = \frac{1}{(1+\chi)g_m} \tag{8.95}$$

EXERCISE

8.25 For the source follower in Fig. 8.35(a), let the bias current of Q_1 be 200 μA and assume Q_1 is operating at $V_{OV} = 0.2$ V. If $V_A = 5$ V and $\chi = 0.2$, find the voltage gain and output resistance of the source follower.
Ans. 0.81 V/V; 403 Ω

8.7 Current-Mirror Circuits with Improved Performance



As we have seen throughout this chapter, current sources play a major role in the design of IC amplifiers. The constant-current source is used both in biasing and as an active load, and in Section 8.2 we studied simple forms of both MOS and bipolar current sources and, more generally, current mirrors. We also demonstrated the need to improve the characteristics of simple sources and mirrors. Specifically, three performance parameters need to be addressed:

1. The accuracy of the current transfer ratio of the mirror. For bipolar mirrors, this parameter is primarily affected by the transistor β. For both bipolar and MOS mirrors, the Early effect affects the current transfer ratio.
2. The output resistance, R_o. Among other reasons, the need to increase the output resistance of current sources stems from the need to increase the voltage gain achievable in an amplifier stage. While simple bipolar and MOS mirrors have output resistances equal to r_o, cascoding can be used to increase the output resistance.
3. The minimum dc voltage required across the current source. The need to keep this voltage as small as possible stems from the low dc voltage supplies used in modern IC technologies. Simple BJT and MOS sources can operate with dc voltages in the range of 0.2 V to 0.3 V. More elaborate mirror circuits, unfortunately, tend to require higher voltages.

In this section we study MOS and bipolar current mirrors that feature improvements in one or more of these characteristics.

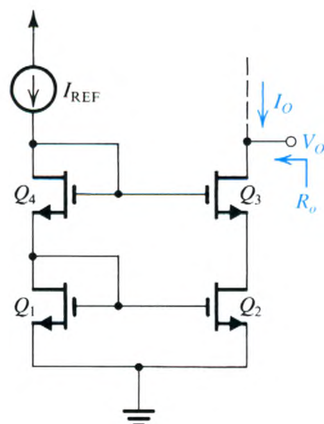


Figure 8.36 A cascode MOS current mirror.

8.7.1 The Cascode MOS Mirror

The use of cascoding in the design of current sources was presented in Section 8.5. Figure 8.36 shows the basic cascode current mirror. Observe that in addition to the diode-connected transistor Q_1, which is part of the basic mirror Q_1–Q_2, another diode-connected transistor, Q_4, is used to provide a suitable bias voltage for the gate of the cascode transistor Q_3. To determine the output resistance of the cascode mirror at the drain of Q_3, we assume that because of the low incremental resistances of the diode-connected transistors Q_1 and Q_4, the voltages across them will remain almost constant, and thus the signal voltages at the gates of Q_2 and Q_3 will be nearly zero. Thus R_o will be that of the cascode current source formed by Q_2 and Q_3,

$$R_o \simeq g_{m3} r_{o3} r_{o2} \tag{8.96}$$

Thus, as expected, cascoding raises the output resistance of the current mirror by the factor $(g_{m3} r_{o3})$, which is the intrinsic gain of the cascode transistor.

A drawback of the cascode current mirror is that it consumes a relatively large portion of the steadily shrinking supply voltage V_{DD}. While the simple MOS mirror operates properly with a voltage as low as V_{OV} across its output transistor, the cascode circuit of Fig. 8.36 required a minimum voltage of $V_t + 2V_{OV}$. This is because the gate of Q_3 is at $2V_{GS} = 2V_t + 2V_{OV}$ and the voltage at the drain of Q_3 can go below that at the gate by a maximum of V_t. Thus the minimum voltage required across the output of the cascode mirror is 1 V or so. This obviously limits the signal swing at the output of the mirror (i.e., at the output of the amplifier that utilizes this current source as a load). In Chapter 13, we will study a modified cascode mirror known as the *wide-swing current mirror*.

EXERCISE

8.26 For a cascode MOS mirror utilizing devices with $V_t = 0.5$ V, $\mu_n C_{ox} = 387\ \mu\text{A/V}^2$, $V_A' = 5$ V/μm, $W/L = 3.6\ \mu\text{m}/0.36\ \mu\text{m}$, and $I_{REF} = 100\ \mu\text{A}$, find the minimum voltage required at the output and the output resistance.
Ans. 0.95 V; 285 kΩ

8.7.2 The Wilson BJT Current Mirror

A simple but ingenious modification of the basic bipolar mirror results in both reducing the β dependence and increasing the output resistance. The resulting circuit, known as the **Wilson mirror** after its inventor George Wilson, an IC design engineer working at the time

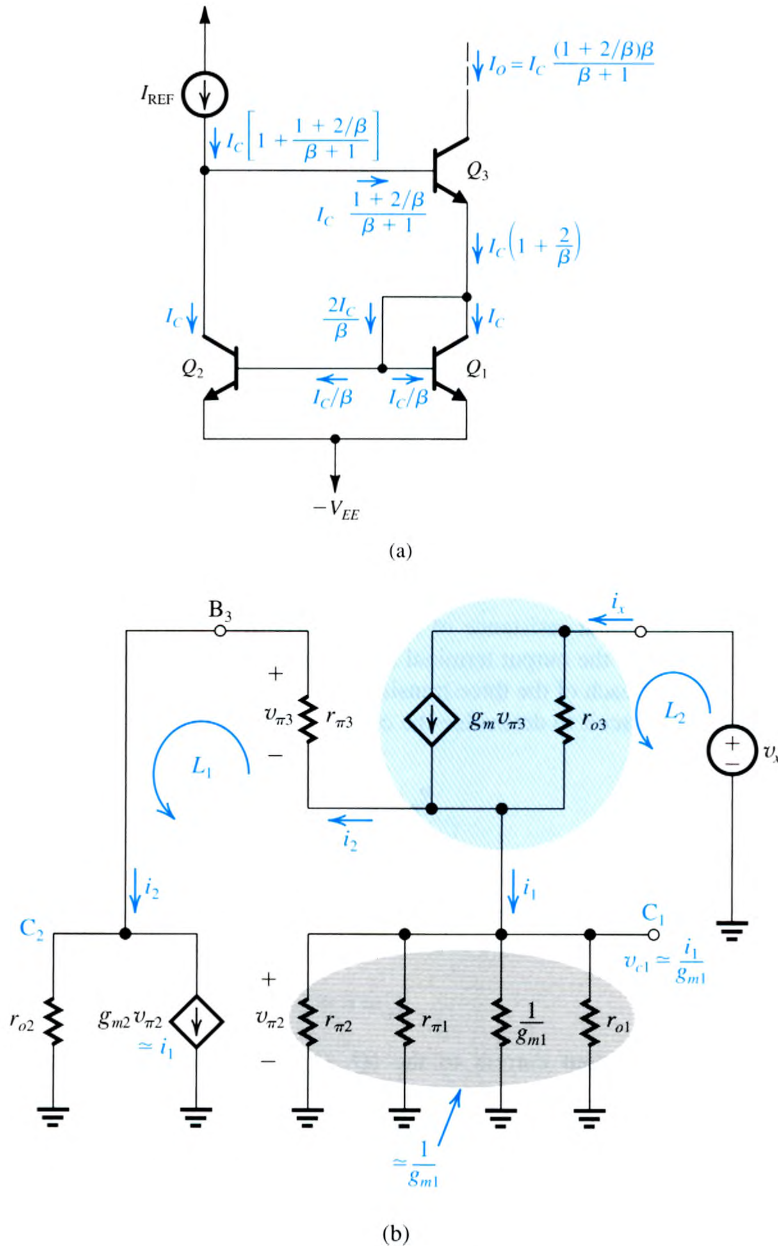


Figure 8.37 The Wilson bipolar current mirror: (**a**) circuit showing analysis to determine the current transfer ratio; (**b**) determining the output resistance.

for Tektronix, is shown in Fig. 8.37(a). Note that the circuit employs negative feedback: A current almost equal to the output current I_O is fed to the basic current mirror formed by Q_1 and Q_2 with the output current of this mirror, in the collector of Q_2, fed back to the input node of the mirror. The analysis to determine the effect of finite β on the current transfer ratio is shown in Fig. 8.37(a), from which we can write

$$\frac{I_O}{I_{\text{REF}}} = \frac{I_C\left(1+\frac{2}{\beta}\right)\beta\Big/(\beta+1)}{I_C\left[1+\left(1+\frac{2}{\beta}\right)\Big/(\beta+1)\right]}$$

$$= \frac{\beta+2}{\beta+1+\frac{\beta+2}{\beta}} = \frac{\beta+2}{\beta+2+\frac{2}{\beta}}$$

$$= \frac{1}{1+\frac{2}{\beta(\beta+2)}}$$

$$\simeq \frac{1}{1+2/\beta^2} \tag{8.97}$$

which is much less dependent on β than in the case of the simple current mirror.

This analysis assumes that Q_1 and Q_2 conduct equal collector currents. There is, however, a slight problem with this assumption: The collector-to-emitter voltages of Q_1 and Q_2 are not equal, which introduces a current offset or a systematic error. This problem can be solved by adding a diode-connected transistor in series with the collector of Q_2, as we shall shortly show for the MOS version.

To determine the output resistance of the Wilson current source, we set $I_{\text{REF}} = 0$ and apply a test voltage v_x to the output terminal (the collector of Q_3) as shown in Fig. 8.37(b). Here we have replaced each of the three transistors with its hybrid-π equivalent circuit. Our goal is to analyze the circuit to determine the current i_x drawn from v_x and hence the output resistance R_o,

$$R_o \equiv \frac{v_x}{i_x}$$

The analysis proceeds as follows:

1. The current i_x enters the supernode formed by the controlled source $g_{m3}v_{\pi3}$ and resistance r_{o3} and divides into two currents, i_1 and i_2, that exit the supernode. Thus,

$$i_x = i_1 + i_2$$

2. Current i_1, the input current to the (Q_1, Q_2) mirror, flows into the total resistance between node C_1 and ground, which is dominated by the resistance of diode-connected transistor Q_1 and thus is approximately equal to $1/g_{m1}$. Thus,

$$v_{\pi2} = v_{c1} \simeq i_1/g_{m1}$$

3. The controlled current-source of Q_2 is

$$g_{m2}v_{\pi2} \simeq g_{m2}\frac{i_1}{g_{m1}} = i_1$$

4. A loop equation for loop L_1 gives

$$v_{c1} = i_2 r_{\pi 3} + (i_2 - i_1) r_{o2}$$

Substituting for $v_{c1} = i_1/g_{m1}$ gives us the following relationship between i_2 and i_1:

$$\frac{i_2}{i_1} = \frac{r_{o2} + \dfrac{1}{g_{m1}}}{r_{o2} + r_{\pi 3}} \simeq 1$$

That is, as expected, the current mirror (Q_1, Q_2) provides an output current i_2 that is approximately equal to i_1. Since $i_1 + i_2 = i_x$, then

$$i_1 = i_2 = \frac{i_x}{2}$$

5. Finally, we write a loop equation for L_2:

$$v_x = (i_x - g_{m3} v_{\pi 3}) r_{o3} + \frac{i_1}{g_{m1}}$$

Substituting $v_{\pi 3} = -i_2 r_{\pi 3} = -\left(\dfrac{i_x}{2}\right) r_{\pi 3}$ and $i_1 = \dfrac{i_x}{2}$, and rearranging gives us

$$R_o \equiv \frac{v_x}{i_x} = \left(\frac{1}{2}\beta_3 + 1\right) r_{o3} + \frac{1}{2g_{m1}}$$

Since

$$\frac{1}{2}\beta_3 \gg 1 \text{ and } \beta_3 r_{o3} \gg \frac{1}{g_{m1}},$$

$$R_o \simeq \frac{1}{2}\beta_3 r_{o3} \tag{8.98}$$

Thus the Wilson current mirror has an output resistance $(\frac{1}{2}\beta_3)$ times higher than that of Q_3 alone. This is a result of the negative feedback obtained by feeding the collector current of Q_2 (i_2) back to the base of Q_3. As we can see from the above analysis, this feedback increases the current through r_{o3} to approximately $\frac{1}{2}\beta_3 i_x$, and thus the voltage across r_{o3} and the output resistance increase by the same factor, $\frac{1}{2}\beta_3$. Finally, note that the factor $\frac{1}{2}$ comes about because only half of i_x is mirrored back to the base of Q_3.

The Wilson mirror is preferred over the cascode circuit because the latter has the same dependence on β as the simple mirror. However, like the cascode mirror, the Wilson mirror requires an additional V_{BE} drop for its operation; that is, for proper operation we must allow for 1 V or so across the Wilson mirror output.

EXERCISE

8.27 For $\beta = 100$ and $r_o = 100\,\text{k}\Omega$, contrast the Wilson mirror and the simple mirror by evaluating the current transfer-ratio error due to finite β, and the output resistance.
Ans. Transfer-ratio error: 0.02% for Wilson as opposed to 2% for the simple circuit; $R_o = 5\,\text{M}\Omega$ for Wilson compared to $100\,\text{k}\Omega$ for the simple circuit

8.7.3 The Wilson MOS Mirror

Figure 8.38(a) shows the MOS version of the Wilson mirror. Obviously, there is no β error to reduce here, and the advantage of the MOS Wilson lies in its enhanced output resistance. However, before delving into the analysis, we should take a moment to appreciate how the circuit works: The output current I_O in Q_3 is fed to the input of the basic current mirror formed by Q_1 and Q_2. This mirror in turn provides in the drain of Q_2 a current that, neglecting for the moment the Early effect, is equal to I_O. This latter current is *fed back* to the input node, where it is forced to equal I_{REF} (because the gate current of Q_3 is zero). This negative feedback is the essential feature of the Wilson mirror, giving it its high output resistance, as we will now show.

To determine the output resistance R_o, we set $I_{\text{REF}} = 0$ and apply a test voltage v_x to the output terminal (the drain of Q_3), as shown in Fig. 8.38(b). Here we have replaced each of the three transistors with its hybrid-π equivalent circuit. Our goal is to analyze the circuit to determine the current i_x drawn from v_x and hence the output resistance R_o,

$$R_o \equiv \frac{v_x}{i_x}$$

The analysis proceeds as follows:

1. The current i_x enters the supernode formed by the controlled source $g_{m3}v_{gs3}$ and resistance r_{o3}, and exits the supernode at node S_3 and flows into the diode-connected transistor Q_1.
2. The diode-connected transistor Q_1 has a resistance approximately equal to $1/g_{m1}$. The current i_x, which is the input current to the mirror (Q_1, Q_2), establishes a voltage across Q_1 of approximately i_x/g_{m1}. Thus

$$v_{gs2} \simeq i_x/g_{m1}$$

3. The controlled current-source of Q_2 is

$$g_{m2}v_{gs2} = g_{m2}\frac{i_x}{g_{m1}} = i_x$$

which is the output current of the mirror (Q_1, Q_2).

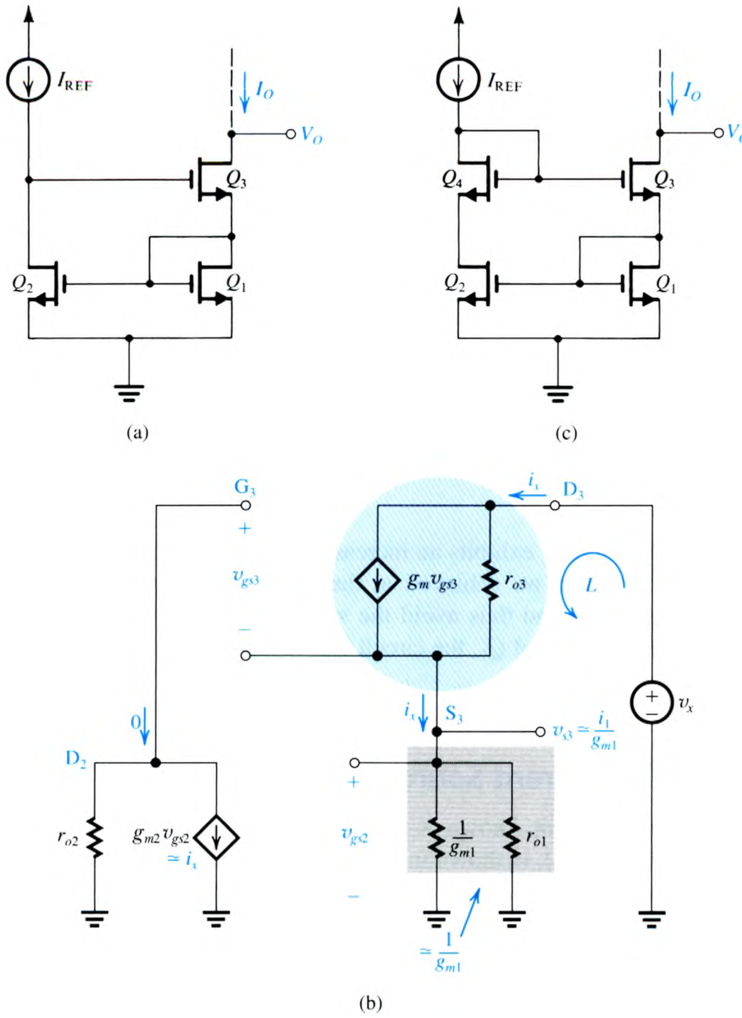


Figure 8.38 The Wilson MOS mirror: (**a**) circuit; (**b**) analysis to determine output resistance; (**c**) modified circuit.

4. Since the input current of Q_3 is zero, the output current of Q_2 flows into r_{o2}, with the result that

$$v_{g3} = v_{d2} = -i_x r_{o2}$$

Thus the voltage v_{gs3} is given by

$$v_{gs3} = v_{g3} - v_{s3} = -i_x r_{o2} - \frac{i_x}{g_{m1}} \simeq -i_x r_{o2}$$

and the controlled current-source of Q_3 is

$$g_{m3} v_{gs3} = -(g_{m3} r_{o2}) i_x$$

5. Finally, a loop equation around L gives us

$$v_x = (i_x + g_{m3} r_{o2} i_x) r_{o3} + \frac{i_x}{g_{m1}}$$

which can be rearranged to yield

$$R_o \equiv \frac{v_x}{i_x} = (g_{m3} r_{o3}) r_{o2} + r_{o3} + \frac{1}{g_{m1}}$$

which can be approximated as

$$R_o \simeq (g_{m3} r_{o3}) r_{o2} \tag{8.99}$$

Thus, the Wilson MOS mirror exhibits an increase of output resistance by a factor $(g_{m3} r_{o3})$, a result identical to the one we achieved in the cascode mirror. Finally, to balance the two branches of the mirror and thus avoid the systematic current error resulting from the difference in V_{DS} between Q_1 and Q_2, the circuit can be modified as shown in Fig. 8.38(c).

8.7.4 The Widlar Current Source

Our final current-source circuit, known as the **Widlar current source** (after Robert Widlar, a pioneer in analog IC design), is shown in Fig. 8.39. It differs from the basic current-mirror circuit in an important way: A resistor R_E is included in the emitter lead of Q_2. Neglecting base currents, we can write

$$V_{BE1} = V_T \ln\left(\frac{I_{\text{REF}}}{I_S}\right) \tag{8.100}$$

and

$$V_{BE2} = V_T \ln\left(\frac{I_O}{I_S}\right) \tag{8.101}$$

where we have assumed that Q_1 and Q_2 are matched devices. Combining Eqs. (8.100) and (8.101) gives us

$$V_{BE1} - V_{BE2} = V_T \ln\left(\frac{I_{\text{REF}}}{I_O}\right) \tag{8.102}$$

But from the circuit we see that

$$V_{BE1} = V_{BE2} + I_O R_E \tag{8.103}$$

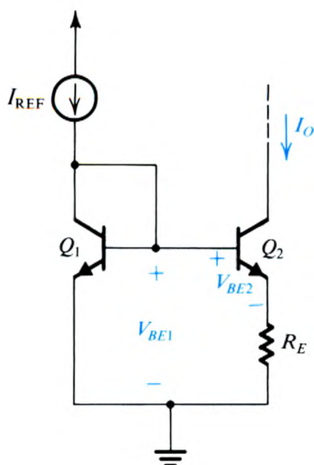


Figure 8.39 The Widlar current source.

Thus,

$$I_O R_E = V_T \ln\left(\frac{I_{REF}}{I_O}\right) \tag{8.104}$$

The design and advantages of the Widlar current source are illustrated in the following example.

Example 8.5

The two circuits for generating a constant current $I_O = 10\ \mu\text{A}$ shown in Fig. 8.40 operate from a 3-V supply. Determine the required values of the resistors, assuming that V_{BE} is 0.7 V at a current of 1 mA and neglecting the effect of finite β.

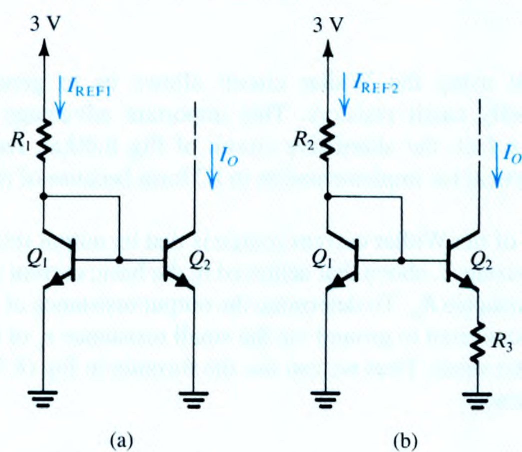


Figure 8.40 Circuits for Example 8.5.

Example 8.5 *continued*

Solution

For the basic current-source circuit in Fig. 8.40(a) we choose a value for R_1 to result in $I_{REF1} = 10\ \mu A$. At this current, the voltage drop across Q_1 will be

$$V_{BE1} = 0.7 + V_T \ln\left(\frac{10\ \mu A}{1\ mA}\right) = 0.58\ V$$

Thus,

$$R_1 = \frac{3 - 0.58}{0.01} = 242\ k\Omega$$

For the Widlar circuit in Fig. 8.40(b) we must first decide on a suitable value for I_{REF2}. If we select $I_{REF2} = 1$ mA, then $V_{BE1} = 0.7$ V and R_2 is given by

$$R_2 = \frac{3 - 0.7}{1} = 2.3\ k\Omega$$

The value of R_3 can be determined using Eq. (8.104) as follows:

$$10 \times 10^{-6} R_3 = 0.025 \ln\left(\frac{1\ mA}{10\ \mu A}\right)$$

$$R_3 = 11.5\ k\Omega$$

The above example shows that using the Widlar circuit allows us to generate a small constant current using relatively small resistors. This important advantage yields considerable savings in chip area. In fact, the alternative circuit of Fig. 8.40(a), requiring a 242-kΩ resistance, is totally impractical for implementation in IC form because of the very high value of resistor R_1.

Another important characteristic of the Widlar current source is that its output resistance is high. The increase in the output resistance, above that achieved in the basic current source, is due to the emitter-degeneration resistance R_E. To determine the output resistance of Q_2, we assume that since the base of Q_2 is connected to ground via the small resistance r_e of Q_1, the incremental voltage at the base will be small. Thus we can use the formula in Eq. (8.72) and adapt it for our purposes here as follows:

$$R_{out} \simeq \left[1 + g_m (R_E \| r_\pi)\right] r_o \tag{8.105}$$

Thus the output resistance is increased above r_o by a factor that can be significant.

EXERCISE

8.28 Find the output resistance of each of the two current sources designed in Example 8.5. Let $V_A = 100$ V and $\beta = 100$.

Ans. 10 MΩ; 54 MΩ

Summary

- Integrated-circuit fabrication technology offers the circuit designer many exciting opportunities, the most important of which is the large number of inexpensive small-area MOS transistors. An overriding concern for IC designers, however, is minimizing chip area or "silicon real estate." As a result, large-valued resistors and capacitors are virtually absent.
- Biasing in integrated circuits utilizes current sources. As well, current sources are used as load devices. Typically, an accurate and stable reference current is generated and then replicated to provide bias currents for the various amplifier stages on the chip. The heart of the current-steering circuitry utilized to perform this function is the current mirror.
- The MOS current mirror has a current transfer ratio of $(W/L)_2/(W/L)_1$. For a bipolar mirror, the ratio is I_{S2}/I_{S1}.
- Bipolar mirrors suffer from the finite β, which reduces the accuracy of the current transfer ratio.
- Both bipolar and MOS mirrors of the basic type have a finite output resistance equal to r_o of the output device. Also, for proper operation, a voltage of at least 0.3 V is required across the output transistor of a simple bipolar mirror ($|V_{OV}|$ for the MOS case).
- The basic gain cell of IC amplifiers is the CS (CE) amplifier with a current-source load. For an ideal current-source load (i.e., one with infinite output resistance), the transistor operates in an open-circuit fashion and thus provides the maximum gain possible, $A_{vo} = -g_m r_o = -A_0$.
- The intrinsic gain A_0 is given by $A_0 = V_A/V_T$ for a BJT and $A_0 = V_A/(V_{OV}/2)$ for a MOSFET. For a BJT, A_0 is constant independent of bias current and device dimensions. For a MOSFET, A_0 is inversely proportional to $\sqrt{I_D}$ (see Eq. 8.46).
- Simple current-source loads reduce the gain realized in the basic gain cell because of their finite output resistance (usually comparable to the value of r_o of the amplifying transistor).
- To raise the output resistance of the CS or CE transistor, we stack a CG or CB transistor on top. This is cascoding.
- The CG and CB amplifiers act as current buffers. They have a short-circuit current gain of unity. For the CG: $R_{in} = \frac{r_o+R_L}{g_m r_o}$ and $R_{out} = R_s + r_o + g_m r_o R_s$. For the CB: $R_{in} = r_\pi \parallel \left[\frac{r_o+R_L}{g_m r_o}\right]$ and $R_{out} = (R_e \parallel r_\pi) + r_o + g_m r_o (R_e \parallel r_\pi)$.
- The CG or CB transistor in the cascode passes the current $g_{m1} v_i$ provided by the CS or CE transistor to the output but increases the resistance at the output from r_{o1} to $(g_{m2} r_{o2}) r_{o1}$ in the MOS case $[g_{m2} r_{o2} (r_{o1} \parallel r_{\pi 2})$ in the bipolar case]. The maximum output resistance achieved in the bipolar case is $\beta_2 r_{o2}$.
- A MOS cascode amplifier operating with an ideal current-source load achieves a gain of $(g_m r_o)^2 = A_0^2$.
- To realize the full advantage of cascoding, the load current-source must also be cascoded, in which case a gain as high as $\frac{1}{2} A_0^2$ can be obtained.
- A CS amplifier with a resistance R_s in its source lead has an output resistance $R_o \simeq (1 + g_m R_s) r_o$. The corresponding formula for the BJT case is $R_o = [1 + g_m (R_e \parallel r_\pi)] r_o$.
- Cascoding can be applied to current mirrors to increase their output resistances. An alternative that also solves the β problem in the bipolar case is the Wilson circuit. The MOS Wilson mirror has an output resistance of $(g_m r_o) r_o$, and the BJT version has an output resistance of $\frac{1}{2} \beta r_o$. Both the cascode and Wilson mirrors require at least 1 V or so for proper operation.
- The Widlar current source provides an area-efficient way to implement a low-valued constant-current source that also has a high output resistance.

Computer Simulation Problems

SIM Problems identified by this icon are intended to demonstrate the value of using SPICE simulation to verify hand analysis and design, and to investigate important issues such as allowable signal swing and amplifier nonlinear distortion. Instructions to assist in setting up simulations for all the indicated problems can be found in the corresponding files on the website. Note that if a particular parameter value is not specified in the problem statement, you are to make a reasonable assumption.

Section 8.2: IC Biasing—Current Sources and Current Mirrors

D 8.1 Using two matched MOS transistors with $W/L = 10$, $k_n' = 400\ \mu\text{A/V}^2$, and $V_{tn} = 0.5$ V, design the circuit in Fig. 8.1 to provide $I_O = 80\ \mu\text{A}$. Assume $V_{DD} = 1.8$ V and neglect the effect of channel-length modulation. Specify the value required for R and the minimum value that V_O can have while Q_2 still operates in saturation.

VE 8.1 **8.2** For $V_{DD} = 1.2$ V and using $I_{REF} = 10\ \mu\text{A}$, it is required to design the circuit of Fig. 8.1 to obtain an output current whose nominal value is 60 µA. Find R and W_2 if Q_1 and Q_2 have equal channel lengths of 0.4 µm, $W_1 = 1$ µm, $V_t = 0.4$ V, and $k_n' = 400\ \mu\text{A/V}^2$. What is the lowest possible value of V_O? Assuming that for this process technology, the Early voltage $V_A' = 6$ V/µm, find the output resistance of the current source. Also, find the change in output current resulting from a +0.2-V change in V_O.

D 8.3 Using $V_{DD} = 1.8$ V and a pair of matched MOSFETs, design the current-source circuit of Fig. 8.1 to provide an output current of 150-µA nominal value. To simplify matters, assume that the nominal value of the output current is obtained at $V_O \simeq V_{GS}$. The circuit must operate for V_O in the range of 0.3 V to V_{DD} and the change in I_O over this range must be limited to 10% of the nominal value of I_O. Find the required value of R and the device dimensions. For the fabrication-process technology used, $\mu_n C_{ox} = 400\ \mu\text{A/V}^2$, $V_A' = 10$ V/µm, and $V_t = 0.5$ V.

D 8.4 Sketch the p-channel counterpart of the current-source circuit of Fig. 8.1. Let the sources of Q_1 and Q_2 be connected to V_{DD}. Let $V_{DD} = 1.3$ V, $|V_t| = 0.4$ V, Q_1 and Q_2 be matched, and $\mu_p C_{ox} = 80\ \mu\text{A/V}^2$. Find the device W/L ratios and the value of the resistor that sets the value of I_{REF} so that a nominally 50-µA output current is obtained. The current source is required to operate for V_O as high as 1.1 V. Neglect channel-length modulation.

SIM **D 8.5** A current source is implemented using the basic NMOS current mirror with $I_{REF} = 25\ \mu\text{A}$, $L_1 = L_2 = 0.54$ µm, $W_1 = 1.74$ µm, and $W_2 = 4W_1$. The circuit is fabricated in the 0.18-µm CMOS technology with the parameters given in Table K.1 (Appendix K). Determine V_{OV}, $I_{Onominal}$, V_{CSmin}, and R_o. Also, find I_O for $V_O = 0.2$ V, 0.7 V, 1.2 V, and 1.7 V.

8.6 For the current-steering circuit of Fig. P8.6, find I_O in terms of I_{REF} and device W/L ratios.

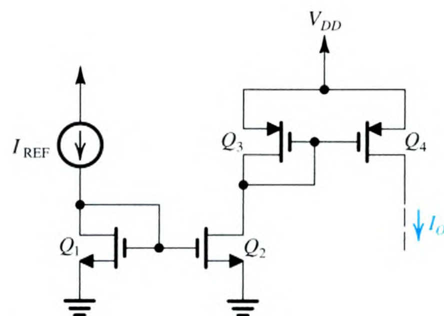


Figure P8.6

D 8.7 The current-steering circuit of Fig. P8.7 is fabricated in a CMOS technology for which $\mu_n C_{ox} = 400\ \mu\text{A/V}^2$, $\mu_p C_{ox} = 100\ \mu\text{A/V}^2$, $V_{tn} = 0.5$ V, $V_{tp} = -0.5$ V, $V_{An}' = 6$ V/µm, and $|V_{Ap}'| = 6$ V/µm. If all devices have $L = 0.5$ µm, design the circuit so that $I_{REF} = 20\ \mu\text{A}$, $I_2 = 80\ \mu\text{A}$, $I_3 = I_4 = 50\ \mu\text{A}$, and $I_5 = 100\ \mu\text{A}$. Use the minimum possible device widths needed to operate the current source Q_2 with voltages at its drain as high as +0.8 V and to operate the current sink Q_5 with voltages at its drain as low as –0.8 V. Specify the widths of all devices and the value of R. Find the output resistance of the current source Q_2 and the output resistance of the current sink Q_5.

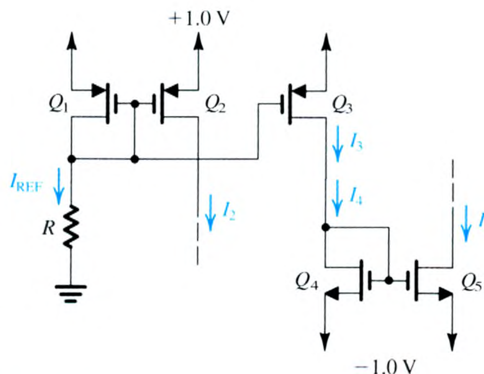


Figure P8.7

***8.8** A PMOS current mirror consists of three PMOS transistors, one diode connected and two used as current outputs. All transistors have $|V_t| = 0.6$ V, $k_p' = 100$ μA/V^2, and $L = 1.0$ μm but three different widths, namely, 10 μm, 20 μm, and 40 μm. When the diode-connected transistor is supplied from a 100-μA source, how many different output currents are available? Repeat with two of the transistors, diodes connected and the third used to provide current output. For each possible input-diode combination, give the values of the output currents and of the V_{SG} that results.

8.9 Consider the BJT current mirror of Fig. 8.7 with a nominal current transfer ratio of unity. Let the transistors have $I_S = 6 \times 10^{-18}$ A, $\beta = 100$, and $V_A = 35$ V. For $I_{REF} = 100$ μA, find I_O when $V_O = 2$ V. Also, find the output resistance and the change in I_O resulting from V_O changing by 1.0 V.

8.10 Consider the basic BJT current mirror of Fig. 8.7 when Q_1 and Q_2 are matched and $I_{REF} = 0.2$ mA. Neglecting the effect of finite β, find the change in I_O, both as an absolute value and as a percentage, corresponding to V_O changing from 0.5 V to 3 V. The Early voltage is 50 V.

8.11 Consider the basic bipolar current mirror of Fig. 8.7 for the case in which Q_1 and Q_2 are identical devices having $I_S = 10^{-17}$ A.

(a) Assuming the transistor β is very high, find the range of V_{BE} and I_O corresponding to I_{REF} increasing from 10 μA to 10 mA. Assume that Q_2 remains in the active mode, and neglect the Early effect.

(b) Find the range of I_O corresponding to I_{REF} in the range of 10 μA to 10 mA, taking into account the finite β. Assume that β remains constant at 100 over the current range 0.1 mA to 5 mA but that at $I_C \simeq 10$ mA and at $I_C \simeq 10$ μA, $\beta = 50$. Specify I_O corresponding to $I_{REF} =$ 10 μA, 0.1 mA, 1 mA, and 10 mA. Note that β variation with current causes the current transfer ratio to vary.

8.12 Consider the basic BJT current mirror of Fig. 8.7 for the case in which Q_2 has m times the area of Q_1. Show that the current transfer ratio is given by Eq. (8.19). If β is specified to be a minimum of 50, what is the largest current transfer ratio m possible if the error introduced by the finite β is limited to 10%?

D 8.13 Give the circuit for the *pnp* version of the basic current mirror of Fig. 8.7. If β of the *pnp* transistor is 50, what is the current gain (or transfer ratio) I_O/I_{REF} for the case of identical transistors, neglecting the Early effect?

D 8.14 The current-source circuit of Fig. P8.14 uses a pair of matched *pnp* transistors having $I_S = 10^{-15}$ A, $\beta = 50$, and $|V_A| = 50$ V. Design the circuit to provide an output current $I_O = 1$ mA at $V_O = 1$ V. What values of I_{REF} and R are needed? What is the maximum allowed value of V_O while the current source continues to operate properly? What change occurs in I_O corresponding to V_O changing from the maximum positive value to –5 V? *Hint:* Adapt Eq. (8.21) for this case as:

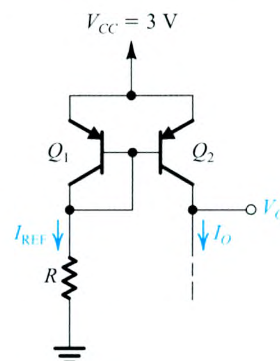


Figure P8.14

$$I_O = I_{REF}\left[\frac{1 + \dfrac{3 - V_O - V_{EB}}{|V_A|}}{1 + \dfrac{2}{\beta}}\right]$$


VE 8.2

8.15 Find the voltages at all nodes and the currents through all branches in the circuit of Fig. P8.15. Assume $|V_{BE}| = 0.7$ V and $\beta = \infty$.

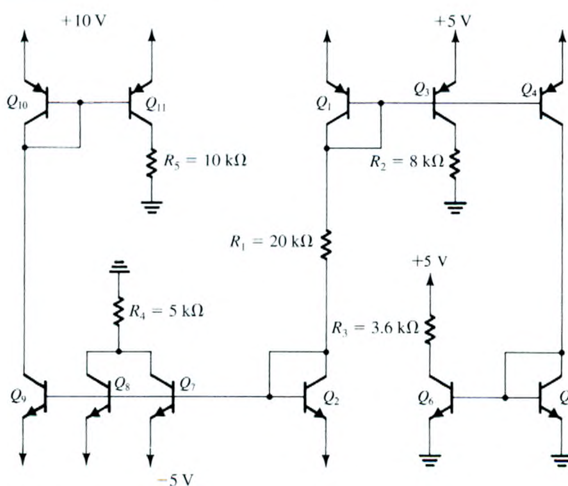


Figure P8.15

8.16 For the circuit in Fig. P8.16, let $|V_{BE}| = 0.7$ V and $\beta = \infty$. Find I, V_1, V_2, V_3, V_4, and V_5 for (a) $R = 10$ kΩ and (b) $R = 100$ kΩ.

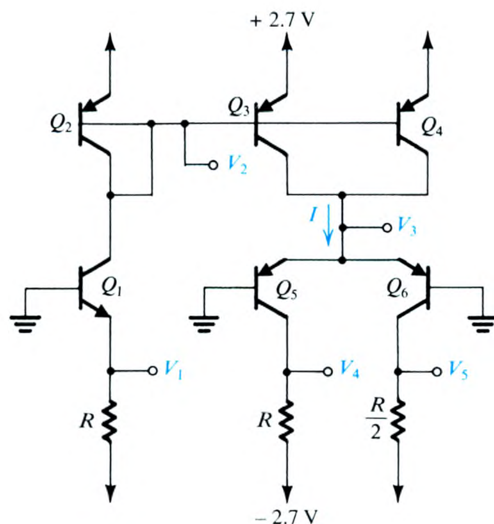


Figure P8.16

D 8.17 Using the ideas embodied in Fig. 8.10, design a multiple-mirror circuit using power supplies of ±3 V to create source currents of 0.1 mA, 0.2 mA, and 0.4 mA and sink currents of 0.3 mA, 0.5 mA, and 1 mA. Assume that the BJTs have $|V_{BE}| \simeq 0.7$ V and large β. What is the total power dissipated in your circuit?

***8.18** The circuit shown in Fig. P8.18 is known as a **current conveyor**.

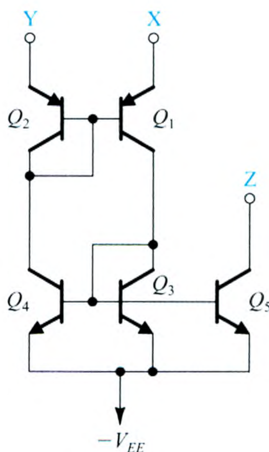


Figure P8.18

(a) Assuming that Y is connected to a voltage V, a current I is forced into X, and terminal Z is connected to a voltage that keeps Q_5 in the active region, show that a current equal to I flows through terminal Y, that a voltage equal to V appears at terminal X, and that a current equal to I flows through terminal Z. Assume β to be large, corresponding transistors are matched, and all transistors are operating in the active region.

(b) With Y connected to ground, show that a virtual ground appears at X. Now, if X is connected to a +5-V supply through a 10-kΩ resistor, what current flows through Z?

8.19 For the base-current-compensated mirror of Fig. 8.11, let the three transistors be matched and specified to have a collector current of 1 mA at $V_{BE} = 0.7$ V. For I_{REF} of 100 μA and assuming $\beta = 100$, what will the voltage at node x be? If I_{REF} is increased to 1 mA, what is the change in V_X? What is the value of I_O obtained with $V_O = V_X$ in both cases? Give the percentage difference between the actual and ideal value of I_O in each case. What is the lowest voltage at the output for which proper current-source operation is maintained?

D 8.20 Extend the current-mirror circuit of Fig. 8.11 to n outputs. What is the resulting current transfer ratio from the input to each output, I_O/I_{REF}? If the deviation from unity is to be kept at 0.2% or less, what is the maximum possible number of outputs for BJTs with $\beta = 100$?

***8.21** For the base-current-compensated mirror of Fig. 8.11, show that the incremental input resistance (seen by the reference current source) is approximately $2V_T/I_{REF}$. Evaluate R_{in} for $I_{REF} = 100$ μA. (*Hint*: Q_3 is operating at a current $I_{E3} = 2I_C/\beta$, where I_C is the operating current of each of Q_1 and Q_2. Replace each transistor with its T model and neglect r_o.)

8.22 The MOSFETs in the current mirror of Fig. 8.12(a) have equal channel lengths of 0.5 μm, $W_1 = 5$ μm, $W_2 = 30$ μm, $\mu_n C_{ox} = 500$ μA/V^2, and $V_A' = 6$ V/μm. If the input bias current is 25 μA, find R_{in}, A_{is}, and R_o.

D 8.23 The MOSFETs in the current mirror of Fig. 8.12(a) have equal channel lengths, $\mu_n C_{ox} = 400$ μA/V^2 and $V_A' = 20$ V/μm. If the input bias current is 200 μA, find W_1, W_2, and L to obtain a nominal short-circuit current gain of 4, an input resistance of 500 Ω, and an output resistance of 20 kΩ. What is the expected error in the current transfer ratio due to channel-length modulation?

8.24 Figure P8.24 shows an amplifier utilizing a current mirror Q_2–Q_3. Here Q_1 is a common-source amplifier fed with $v_I = V_{GS} + v_i$, where V_{GS} is the gate-to-source dc bias

voltage of Q_1 and v_i is a small signal to be amplified. Find the signal component of the output voltage v_O and hence the small-signal voltage gain v_o/v_i. Also, find the small-signal resistance of the diode-connected transistor Q_2 in terms of g_{m2}, and hence the total resistance between the drain of Q_1 and ground. What is the voltage gain of the CS amplifier Q_1? Neglect all r_o's.

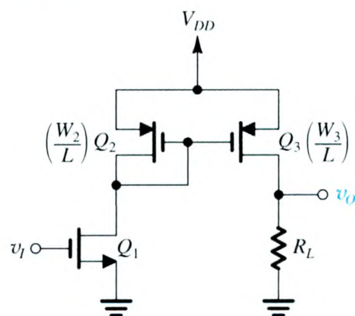


Figure P8.24

***8.25** In a manner parallel to the one used for the MOS current mirror in Section 8.2.5, analyze the BJT mirror circuit in Fig. P8.25 to determine its small-signal input resistance R_{in}, short-circuit current gain $A_{is} \equiv i_o/i_i$, and output resistance R_o. To do this, replace each of Q_1 and Q_2 with its hybrid-π model. Let Q_1 and Q_2 have the same β, but let the EBJ area of Q_2 be m times that of Q_1. Assuming that $r_{o1} \gg 1/g_{m1}$, show that

$$R_{in} \simeq 1/[g_{m1}(1 + \frac{m+1}{\beta})]$$

$$A_{is} = \frac{m}{1 + \frac{m+1}{\beta}}$$

$$R_o = r_{o2}$$

Find the values of R_{in}, A_{is}, and R_o for the case $I = 0.1$ mA, $m = 5$, $\beta = 100$, and $V_A = 30$ V.

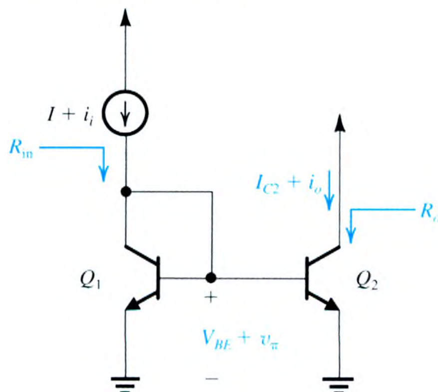


Figure P8.25

8.26 Find the incremental (i.e., small-signal) resistance of each of the diode-connected transistors shown in Fig. P8.26. Assume that the dc bias current $I = 0.2$ mA. For the MOSFET, let $\mu_n C_{ox} = 400\ \mu A/V^2$ and $W/L = 10$. Neglect r_o for both devices.

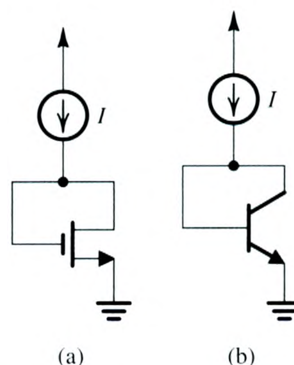


Figure P8.26

Section 8.3: The Basic Gain Cell

8.27 Find g_m, r_π, r_o, and A_0 for the CE amplifier of Fig. 8.13(b) when operated at $I = 20\ \mu A$, $200\ \mu A$, and 2 mA. Assume $\beta = 100$ and remains constant as I is varied, and that $V_A = 16$ V. Present your results in a table.

8.28 Consider the CE amplifiers of Fig. 8.13(b) for the case of $I = 0.5$ mA, $\beta = 200$, and $V_A = 30$ V. Find R_{in}, A_{vo}, and R_o. If you needed to raise R_{in} by a factor of 5 by changing I, what value of I is required, assuming that β remains unchanged? What are the new values of A_{vo} and R_o? If the amplifier is fed with a signal source having $R_{sig} = 10\ k\Omega$ and is connected to a load of 200-kΩ resistance, find the overall voltage gain, v_o/v_{sig}.

8.29 Find the intrinsic gain of an NMOS transistor fabricated in a process for which $k_n' = 400\ \mu A/V^2$ and $V_A' = 10\ V/\mu m$. The transistor has a 0.4-μm channel length and is operated at $V_{OV} = 0.2$ V. If a 1-mA/V transconductance is required, what must I_D and W be?

8.30 An NMOS transistor fabricated in a certain process is found to have an intrinsic gain of 40 V/V when operated at an I_D of 100 μA. Find the intrinsic gain for $I_D = 25\ \mu A$ and $I_D = 400\ \mu A$. For each of these currents, find the factor by which g_m changes from its value at $I_D = 100\ \mu A$.

D 8.31 Consider an NMOS transistor fabricated in a 0.18-μm technology for which $k_n' = 400\ \mu A/V^2$ and $V_A' = 5\ V/\mu m$. It is required to obtain an intrinsic gain of 25 V/V

and a g_m of 1 mA/V. Using $V_{OV} = 0.2$ V, find the required values of L, W/L, and the bias current I.

D 8.32 Sketch the circuit for a current-source-loaded CS amplifier that uses a PMOS transistor for the amplifying device. Assume the availability of a single +1.8-V dc supply. If the transistor is operated with $|V_{OV}| = 0.3$ V, what is the highest instantaneous voltage allowed at the drain?

8.33 An NMOS transistor operated with an overdrive voltage of 0.25 V is required to have a g_m equal to that of an *npn* transistor operated at $I_C = 50$ μA. What must I_D be? What value of g_m is realized?

8.34 For an NMOS transistor with $L = 0.5$ μm fabricated in the 0.13-μm process specified in Table K.1 in Appendix K, find g_m, r_o, and A_0 if the device is operated with $V_{OV} = 0.15$ V and $I_D = 50$ μA. Also, find the required device width W.

8.35 For an NMOS transistor with $L = 0.54$ μm fabricated in the 0.18-μm process specified in Table K.1 in Appendix K, find g_m, r_o, and A_0 obtained when the device is operated at $I_D = 200$ μA with $V_{OV} = 0.2$ V. Also, find W.

8.36 Fill in the table below. For the BJT, let $\beta = 100$ and $V_A = 40$ V. For the MOSFET, let $\mu_n C_{ox} = 400$ μA/ V^2, $W/L = 40$, and $V_A = 10$ V.

	BJT Cell		MOSFET Cell	
Bias Current	$I_C = 1$ mA	$I_C = 0.5$ mA	$I_D = 1$ mA	$I_D = 0.5$ mA
g_m (mA/V)				
r_o (kΩ)				
A_0 (V/V)				
R_{in} (kΩ)				

8.37 A CS amplifier utilizes an NMOS transistor with $L = 0.54$ μm and $W/L = 10$. It was fabricated in a 0.18-μm CMOS process for which $\mu_n C_{ox} = 400$ μA/V^2 and $V_A' = 5$ V/μm. What is the bias current of the transistor for which $A_0 = 20$ V/V?

8.38 A CS amplifier utilizes an NMOS transistor with $L = 0.36$ μm and $W/L = 10$. It was fabricated in a 0.18-μm CMOS process for which $\mu_n C_{ox} = 400$ μA/V^2 and $V_A' = 5$ V/μm. Find the values of g_m and A_0 obtained at $I_D = 20$ μA, 200 μA, and 2 mA.

D 8.39 An NMOS transistor is fabricated in the 0.18-μm process whose parameters are given in Table K.1 in Appendix K. The device has a channel length twice the minimum and is operated at $V_{OV} = 0.25$ V and $I_D = 10$ μA.

(a) What values of g_m, r_o, and A_0 are obtained?
(b) If I_D is increased to 100 μA, what do V_{OV}, g_m, r_o, and A_0 become?
(c) If the device is redesigned with a new value of W so that it operates at $V_{OV} = 0.25$ V for $I_D = 100$ μA, what do g_m, r_o, and A_0 become?
(d) If the redesigned device in (c) is operated at 10 μA, find V_{OV}, g_m, r_o, and A_0.
(e) Which designs and operating conditions produce the lowest and highest values of A_0? What are these values? In each of these two cases, if W/L is held at the same value but L is made 10 times larger, what gains result?

D 8.40 Find A_0 for an NMOS transistor fabricated in a CMOS process for which $k_n' = 400$ μA/V^2 and $V_A' = 6$ V/μm. The transistor has a 0.5-μm channel length and is operated with an overdrive voltage of 0.15 V. What must W be for the NMOS transistor to operate at $I_D = 50$ μA? Also, find the values of g_m and r_o.

D 8.41 The circuit in Fig. 8.15(a) is fabricated in the 0.13-μm CMOS process whose parameters are specified in Table K.1 (Appendix K). The supply voltage $V_{DD} = 1.3$ V. The two transistors have $L = 0.4$ μm and are to be operated at $I_D = 100$ μA and $|V_{OV}| = 0.15$ V. Find the required values of V_G, $(W/L)_1$, $(W/L)_2$, and A_v.

D 8.42 The circuit in Fig. 8.15(a) is fabricated in the 0.18-μm CMOS process whose parameters are specified in Table K.1 (Appendix K). $V_{DD} = 1.8$ V. Design the circuit to obtain a voltage gain $A_v = -20$ V/V. Use devices of equal length L operating at $I = 100$ μA and $|V_{OV}| = 0.2$ V. Determine the required values of V_G, L, $(W/L)_1$, and $(W/L)_2$.

8.43 Figure P8.43 shows an IC MOS amplifier formed by cascading two common-source stages. Assuming that $V_{An} = |V_{Ap}|$ and that the biasing current sources have output resistances equal to those of Q_1 and Q_2, find an expression for the overall voltage gain in terms of g_m and r_o of Q_1 and Q_2. If Q_1 and Q_2 are to be operated at equal overdrive voltages, $|V_{OV}|$, find the required value of $|V_{OV}|$ if $|V_A| = 4$ V and the gain required is 300 V/V.

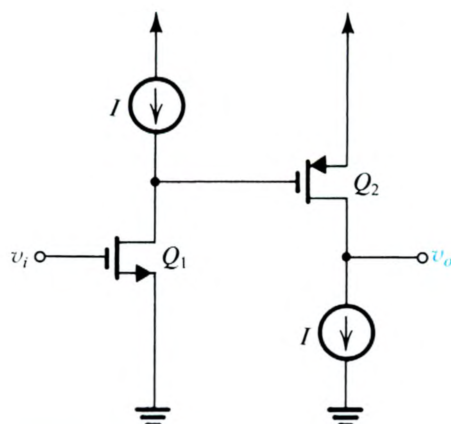


Figure P8.43

***8.44** The NMOS transistor in the circuit of Fig. P8.44 has $V_t = 0.5$ V, $k_n' W/L = 2$ mA/V^2, and $V_A = 20$ V.

(a) Neglecting the dc current in the feedback network and the effect of channel-length modulation, find V_{GS}. Then find the dc current in the feedback network and V_{DS}. Verify that you were justified in neglecting the current in the feedback network when you found V_{GS}.
(b) Find the small-signal voltage gain, v_o/v_i. What is the peak of the largest output sine-wave signal that is possible while the NMOS transistor remains in saturation? What is the corresponding input signal?
(c) Find the small-signal input resistance R_{in}.

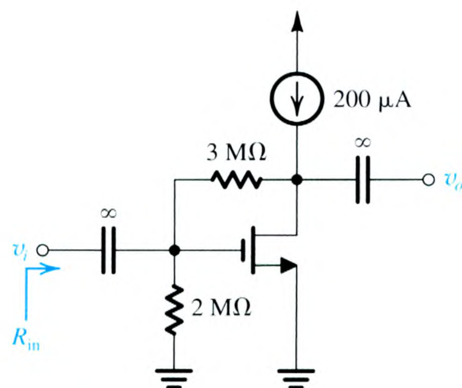


Figure P8.44

SIM 8.45 The CMOS amplifier in Fig.8.16(a) is fabricated in the 0.13-μm CMOS process whose parameters are specified in Table K.1 (Appendix K). All transistors have $L = 0.4$ μm and are operated at $V_{OV} = 0.15$ V, $I_{REF} = 100$ μA and $V_{DD} = 1.3$ V.

(a) Find the dc component of v_I and the W/L ratios of the transistors.
(b) Determine the small-signal voltage gain.
(c) What is the allowable range of signal swing at the output for almost-linear operation?
(d) If the current-source load is replaced with a resistance R_D connected to a power supply V_{DD}' as shown in Fig. 8.16(b), find the values of R_D and V_{DD}' so that I_D, the voltage gain, and the output signal swing remain unchanged.

D 8.46 Consider the CMOS amplifier of Fig. 8.16(a) when fabricated with a process for which $k_n' = 4k_p' = 400$ μA/V^2, $|V_t| = 0.5$ V, and $|V_A| = 5$ V. Find I_{REF} and $(W/L)_1$ to obtain a voltage gain of –20 V/V and an output resistance of 50 kΩ. Recall that Q_2 and Q_3 are matched. If Q_2 and Q_3 are to be operated at the same overdrive voltage as Q_1, what must their W/L ratios be?

D 8.47 Design the CMOS amplifier of Fig. 8.16(a) utilizing the 0.13-μm process whose parameters are specified in Table K.1 (Appendix K). The output voltage must be able to swing to within approximately 0.2 V of the power-supply rails (i.e., from 0.2 V to 1.1 V), and the voltage gain must be about 10 V/V. Design for a dc bias current of 50 μA, and use devices with the same channel length. If the channel length is an integer multiple of the minimum 0.13 μm, what channel length is needed and what W/L ratios are required? What gain is achieved? If it is required to raise the gain by a factor of 2, what channel length would be required, and by what factor does the total gate area of the circuit increase? $V_{DD} = 1.3$ V.

***8.48** The MOSFETs in the circuit of Fig. P8.48 are matched, having $k_n'(W/L)_1 = k_p'(W/L)_2 = 1$ mA/V^2 and $|V_t| = 0.5$ V. The resistance $R = 1$ MΩ.

(a) For G and D open, what are the drain currents I_{D1} and I_{D2}?
(b) For $r_o = \infty$, what is the voltage gain of the amplifier from G to D? (*Hint*: Replace the transistors with their small-signal models.)
(c) For finite r_o($|V_A| = 20$ V), what is the voltage gain from G to D and the input resistance at G?
(d) If G is driven (through a large coupling capacitor) from a source v_{sig} having a resistance of 20 kΩ, find the voltage gain v_d/v_{sig}.
(e) For what range of output signals do Q_1 and Q_2 remain in the saturation region?

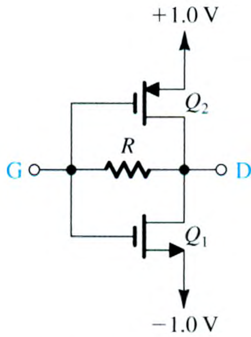


Figure P8.48

8.49 Transistor Q_1 in Fig. P8.49 is operating as a CE amplifier with an active load provided by transistor Q_2, which is the output transistor in a current mirror formed by Q_2 and Q_3. (Note that the biasing arrangement for Q_1 is *not* shown.)

(a) Neglecting the finite base currents of Q_2 and Q_3, and assuming that their $V_{BE} \simeq 0.7$ V and that Q_2 has five times the area of Q_3, find the value of I.

(b) If Q_1 and Q_2 are specified to have $|V_A| = 20$ V, find r_{o1} and r_{o2} and hence the total resistance at the collector of Q_1.

(c) Find $r_{\pi 1}$ and g_{m1} assuming that $\beta_1 = 50$.

(d) Find R_{in}, A_v, and R_o.

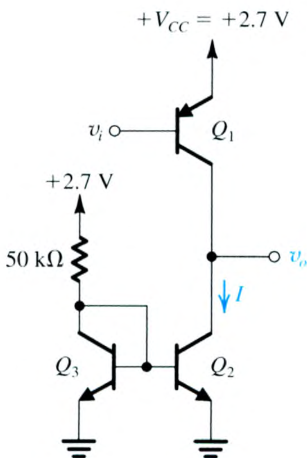


Figure P8.49

Section 8.4: The CG and CB Amplifiers as Current Buffers

8.50 For the CG amplifier in Fig. 8.18, let $g_m = 2$ mA/V, $r_o = 20$ kΩ, $R_s = 20$ kΩ, and $R_L = 840$ kΩ. Determine the values of R_{in}, A_{is}, and R_{out}. If the signal source is converted to its Thévenin form, determine the voltage gain v_o/v_{sig}, where $v_o = i_o R_L$ and $v_{sig} = i_{sig} R_s$.

8.51

(a) For the CG amplifier in Fig. 8.18(a), show that the current gain i_o/i_{sig} is given by

$$\frac{i_o}{i_{sig}} = \frac{R_s}{R_s + R_{in}}$$

(b) For the case $g_m = 2$ mA/V, $r_o = 20$ kΩ, and $R_s = 20$ kΩ, find the current gain obtained with $R_L = 20$ kΩ.

(c) If R_L is doubled (that is, increased by 100%), find the corresponding percentage change in current gain.

(d) If R_L is increased from 20 kΩ to 200 kΩ (that is, increased by 900%), what is the corresponding change in current gain?

(e) Using the results of (c) and (d), comment on the performance of the CG circuit as a current buffer.

D8.52 A CG stage is required to connect a 25-kΩ source to a 1-MΩ load. The CG transistor is to provide an output resistance equal to that of the load. What approximate value of intrinsic gain must the transistor have? If the available process fabrication technology provides $V_A' = 5$ V/μm, specify the required channel length. Operate the transistor at $V_{OV} = 0.15$ V.

D 8.53 Design the current source in Fig. P8.53 to deliver a current of 0.2 mA with an output resistance of 500 kΩ. The transistor has $V_A = 20$ V and $V_t = 0.5$ V. Design for $V_{OV} = 0.2$ V and specify R_s and V_{BIAS}.

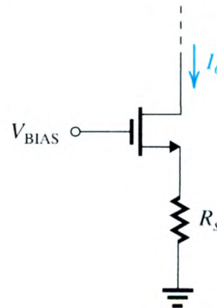


Figure P8.53



D 8.54 Figure P8.54 shows a current source realized using a current mirror with two matched transistors Q_1 and Q_2. Two equal resistances R_s are inserted in the source leads to increase the output resistance of the current source. If Q_2 is operating at $g_m = 2$ mA/V and has $V_A = 5$ V, and if the maximum allowed dc voltage drop across R_s is 0.2 V, what is the maximum available output resistance of the current source? Assume that the voltage at the common-gate node is approximately constant.

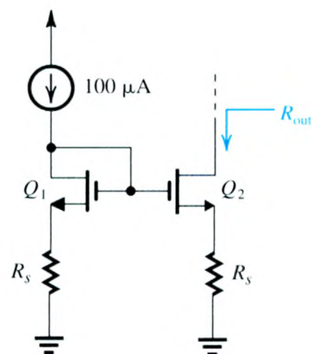


Figure P8.54

8.55 In the common-gate amplifier circuit of Fig. P8.55, Q_2 and Q_3 are matched. $k_n'(W/L)_n = k_p'(W/L)_p = 8$ mA/V^2, and all transistors have $|V_t| = 0.5$ V and $|V_A| = 5$ V.

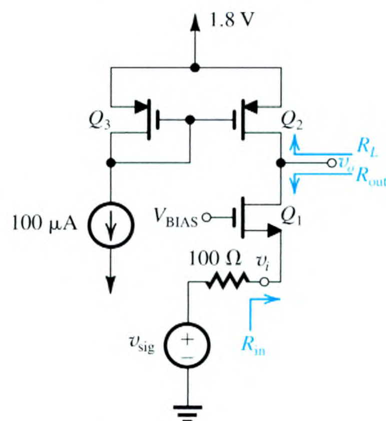


Figure P8.55

The signal v_{sig} is a small sinusoidal signal with no dc component.

(a) Neglecting the effect of V_A, find the dc drain current of Q_1 and the required value of V_{BIAS}.
(b) Find the values of g_{m1} and r_o for all transistors.
(c) Find the value of R_{in}.
(d) Find the value of R_{out}.
(e) Calculate the voltage gains v_o/v_i and v_o/v_{sig}.
(f) How large can v_{sig} be (peak-to-peak) while maintaining saturation-mode operation for Q_1 and Q_2?

8.56 For the CB amplifier in Fig. 8.25, let the transistor be biased at $I_C = 0.1$ mA and have $\beta = 100$ and $V_A = 20$ V. The source resistance $R_e = 25$ kΩ and the load resistance R_L (implemented by a current source) is 1 MΩ. Determine the values of R_{in}, A_{is}, and R_{out}. If the signal source is converted to its Thévenin form, determine the voltage gain v_o/v_{sig}, where $v_o = i_o R_L$ and $v_{sig} = i_{sig} R_e$.

8.57

(a) For the CB amplifier in Fig. 8.25, use the equivalent circuit in Fig. 8.25(b) to obtain an expression for the current gain i_o/i_{sig}.
(b) Find the current gain for the case $g_m = 5$ mA/V, $r_o = 200$ kΩ, $\beta = 100$, $R_e = 20$ kΩ, and $R_L = 200$ kΩ.
(c) If R_L is doubled (that is, increased by 100%), find the corresponding percentage change in current gain.
(d) If R_L is increased from 200 kΩ to 2 MΩ (that is, by 900%), what is the corresponding percentage change in current gain?
(e) Using the results of (c) and (d), comment on the performance of the CB circuit as a current buffer.

8.58 What value of load resistance R_L causes the input resistance of the CB amplifier to be approximately double the value of r_e?

8.59 For the constant-current source circuit shown in Fig. P8.59, find the collector current I and the output resistance. The BJT is specified to have $\beta = 100$, $V_{BE} = 0.7$ V, and $V_A = 25$ V. If the collector voltage undergoes a change of 2 V while the BJT remains in the active mode, what is the corresponding change in collector current?

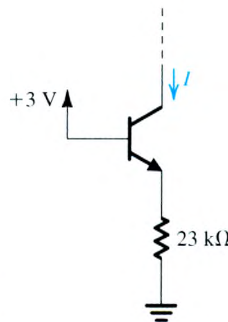


Figure P8.59

D8.60 Consider the constant-current source circuit in Fig. P8.59 for the general case of an emitter resistance R_e. Let $\beta = 100$, $V_A = 25$ V, and $V_{BE} = 0.7$ V. Use the approximate expression for the output resistance R_o,

$$R_o \simeq (g_m r_o)(R_e \| r_\pi)$$

to obtain a relationship between the output resistance R_o and the current I. Hence, find the output resistance for $I = 0.1$ mA, 0.2 mA, 0.5 mA, and 1 mA. For each case, give the required value of R_e.

8.61 Find the value of the resistance R_e, which, when connected in the emitter lead of a CE BJT amplifier, raises the output resistance by a factor of (a) 5, (b) 10, and (c) 50. What is the maximum possible factor by which the output resistance can be raised, and at what value of R_e is it achieved? Assume the BJT has $\beta = 100$ and is biased at $I_C = 0.2$ mA.

Section 8.5: The Cascode Amplifier

D 8.62 In a MOS cascode amplifier, the cascode transistor is required to raise the output resistance by a factor of 20. If the transistor is operated at $V_{OV} = 0.2$ V, what must its V_A be? If the process technology specifies V_A' as 4 V/μm, what channel length must the transistor have?

8.63 The cascode amplifier in Fig. 8.28(a) is fabricated using two identical transistors having $V_A = 3$ V and operated at $V_{OV} = 0.15$ V. What is the realized open-circuit voltage gain?

D 8.64 For a cascode current source such as that in Fig. 8.30, show that if the two transistors are identical, the current I supplied by the current source and the output resistance R_o are related by $IR_o = 2|V_A|^2/|V_{OV}|$. Now consider the case of transistors that have $|V_A| = 3$ V and are operated at $|V_{OV}|$ of 0.15 V. Also, let $\mu_p C_{ox} = 100$ μA/V^2. Find the W/L ratios required and the output resistance realized for the two cases: (a) $I = 50$ μA and (b) $I = 200$ μA. Assume that V_{SD} for the two devices is the minimum required (i.e., $|V_{OV}|$).

D *8.65 For a cascode current source, such as that in Fig. 8.30, show that if the two transistors are identical, the current I supplied by the current source and the output resistance R_o are related by

$$IR_o = \frac{2|V_A'|^2}{|V_{OV}|} L^2$$

Now consider the case of a 0.18-μm technology for which $|V_A'| = 5$ V/μm and let the transistors be operated at $|V_{OV}| =$ 0.2 V. Find the figure-of-merit IR_o for the three cases of L equal to the minimum channel length, twice the minimum, and three times the minimum. Complete the entries of the table at the bottom of the page. Give W/L and the area $2WL$ in terms of n, where n is the value of W/L for the case $I = 0.01$ mA. In the table, A_v denotes the gain obtained in a cascode amplifier such as that in Fig. 8.31 that utilizes our current source as load and which has the same values of g_m and R_o as the current-source transistors.

	$L = L_{min} = 0.18$ μm $IR_o =$ V				$L = 2L_{min} = 0.36$ μm $IR_o =$ V				$L = 3L_{min} = 0.54$ μm $IR_o =$ V			
	g_m (mA/V)	R_o (kΩ)	A_v (V/V)	$2WL$ (μm^2)	g_m (mA/V)	R_o (kΩ)	A_v (V/V)	$2WL$ (μm^2)	g_m (mA/V)	R_o (kΩ)	A_v (V/V)	$2WL$ (μm^2)
$I = 0.01$ mA $W/L = n$												
$I = 0.1$ mA $W/L =$												
$I = 1.0$ mA $W/L =$												

(a) For each current value, what is the price paid for the increase in R_o and A_v obtained as L is increased?
(b) For each value of L, what advantage is obtained as I is increased, and what is the price paid? (*Hint:* We will see in Chapter 10 that the amplifier bandwidth increases with g_m.)
(c) Contrast the performance obtained from the circuit with the largest area with that obtained from the circuit with the smallest area.



VE 8.4

D 8.66 Design the cascode amplifier of Fig. 8.28(a) to obtain $g_{m1} = 3$ mA/V and $R_o = 150$ kΩ. Use the 0.13-μm technology whose parameters are specified in Table K.1 (Appendix K). Determine L, W/L, V_{G2}, and I. Use identical transistors operated at $V_{OV} = 0.15$ V, and design for the maximum possible negative signal swing at the output. What is the value of the minimum permitted output voltage?

D 8.67 Design the circuit of Fig. 8.30 to provide an output current of 100 μA. Use $V_{DD} = 1.3$ V, and assume the PMOS transistors to be identical and have $\mu_p C_{ox} = 128\ \mu\text{A/V}^2$, $V_{tp} = -0.4$ V, and $|V_A| = 3$ V. The current source is to have the widest possible signal swing at its output. Design for $V_{OV} = 0.15$ V, and specify the values of the transistor W/L ratios and of V_{G3} and V_{G4}. What is the highest allowable voltage at the output? What is the value of R_o?

8.68 The cascode amplifier of Fig. 8.31 is operated at a current of 0.2 mA with all devices operating at $|V_{OV}| = 0.25$ V. All devices have $|V_A| = 4$ V. Find g_{m1}, the output resistance of the amplifier, R_{on}, the output resistance of the current source, R_{op}, the overall output resistance, R_o, and the voltage gain, A_v.

D 8.69 Design the CMOS cascode amplifier in Fig. 8.31 for the following specifications: $g_{m1} = 1$ mA/V and $A_v = -280$ V/V. Assume that the amplifier is fabricated in the 0.18-μm CMOS process specified in Table K.1 in Appendix K. Use the same channel length L for all devices and operate all four devices at $|V_{OV}| = 0.2$ V. Determine the required channel length L, the bias current I, and the W/L ratio for each of four transistors. Assume that suitable bias voltages have been chosen, and neglect the Early effect in determining the W/L ratios.

8.70 The cascode transistor can be thought of as providing a "shield" for the input transistor from the voltage variations at the output. To quantify this "shielding" property of the cascode, consider the situation in Fig. P8.70. Here we have grounded the input terminal (i.e., reduced v_i to zero), applied a small change v_x to the output node, and denoted the voltage change that results at the drain of Q_1 by v_y. By what factor is v_y smaller than v_x?

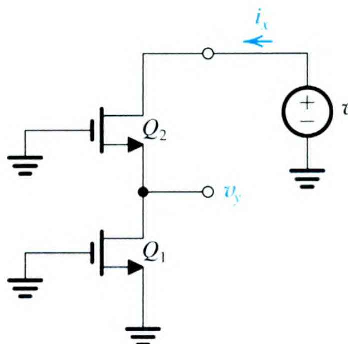


Figure P8.70

8.71 Consider the cascode amplifier of Fig. 8.31 with the dc component at the input $V_I = 0.6$ V, $V_{G2} = 0.9$ V, $V_{G3} = 0.4$ V, $V_{G4} = 0.7$ V, and $V_{DD} = 1.3$ V. If all devices are matched, that is, $k_{n1} = k_{n2} = k_{p3} = k_{p4}$, and have equal $|V_t|$ of 0.4 V, what is the overdrive voltage at which the four transistors are operating? What is the allowable voltage range at the output?

SIM **8.72** A CMOS cascode amplifier such as that in Fig. 8.32(a) has identical CS and CG transistors that have $W/L = 5.4$ μm/0.36 μm and biased at $I = 0.2$ mA. The circuit is fabricated in the 0.18-μm process specified in Table K.1 in Appendix K. At what value of R_L does the gain become −100 V/V? What is the voltage gain of the common-source stage?

8.73 The purpose of this problem is to investigate the signal currents and voltages at various points throughout a cascode amplifier circuit. Knowledge of this signal distribution is very useful in designing the circuit so as to allow for the required signal swings. Figure P8.73 shows a CMOS cascode amplifier with all dc voltages replaced with signal grounds. As well, we have explicitly shown the resistance r_o of each of the four transistors. For simplicity, we are assuming that the four transistors have the same g_m and r_o. The amplifier is fed with a signal v_i.

(a) Determine R_1, R_2, and R_3. Assume $g_m r_o \gg 1$.
(b) Determine i_1, i_2, i_3, i_4, i_5, i_6, and i_7, all in terms of v_i. (*Hint:* Use the current-divider rule at the drain of Q_1.)
(c) Determine v_1, v_2, and v_3, all in terms of v_i.
(d) If v_i is a 5-mV peak sine wave and $g_m r_o = 20$, sketch and clearly label the waveforms of v_1, v_2, and v_3.

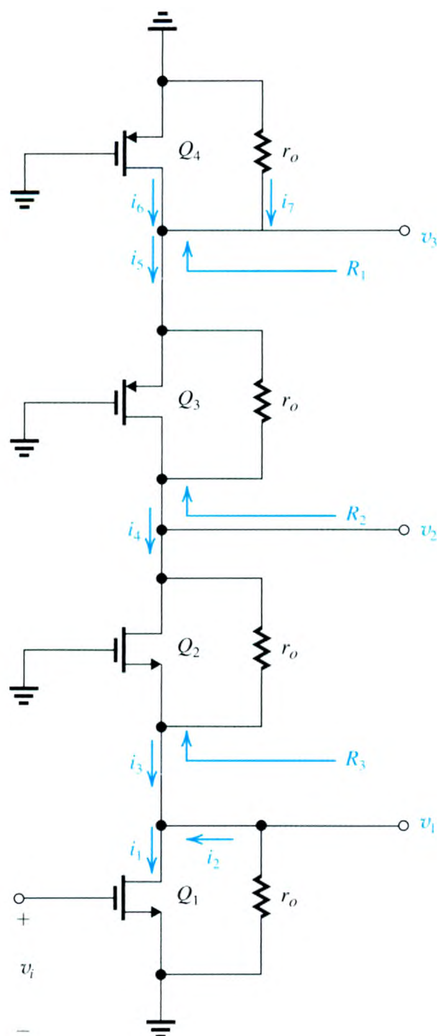


Figure P8.73

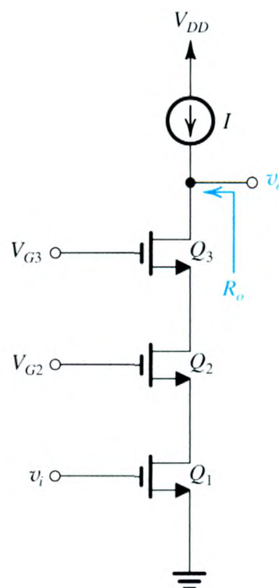


Figure P8.74

8.74 Figure P8.74 shows an amplifier known as a "double cascode." It uses an additional CG transistor Q_3 to increase the resistance by an additional factor $(g_{m3}r_{o3})$. Find an expression for the output resistance R_o and the voltage gain $A_{vo} \equiv v_o/v_i$ (assume the current-source load to be ideal). For identical transistors with $g_m = 2$ mA/V and $r_o = 15$ kΩ, find R_o and A_{vo}.

8.75 A cascode current source formed of two *pnp* transistors for which $\beta = 50$ and $V_A = 6$ V supplies a current of 0.2 mA. What is the output resistance?

8.76 Use Eq. (8.86) to show that for a BJT cascode current source utilizing identical *pnp* transistors and supplying a current I,

$$IR_o = \frac{|V_A|}{(V_T/|V_A|) + (1/\beta)}$$

Evaluate the figure-of-merit IR_o for the case $|V_A| = 5$ V and $\beta = 50$. Now find R_o for the cases of $I = 0.1$, 0.5, and 1.0 mA.

8.77 Consider the BJT cascode amplifier of Fig. 8.34 for the case that all transistors have equal β and r_o. Show that the voltage gain A_v can be expressed in the form

$$A_v = -\frac{1}{2}\frac{|V_A|/V_T}{(V_T/|V_A|) + (1/\beta)}$$

Evaluate A_v for the case $|V_A| = 5$ V and $\beta = 50$. Note that except for the fact that β depends on I as a second-order effect, the gain is independent of the bias current I!

8.78 A bipolar cascode amplifier has a current-source load with an output resistance βr_o. Let $\beta = 50$, $|V_A| = 32$ V, and $I = 0.25$ mA. Find the voltage gain A_v.

8.79 In this problem, we will explore the difference between using a BJT as cascode device and a MOSFET as cascode device. Refer to Fig. P8.79. Given the following data, calculate g_{m1}, R_o, and A_{vo} for the circuits (a) and (b): $I = 100$ µA, $\beta = 125$, $\mu_n C_{ox} = 400$ µA/V^2, $W/L = 25$, $V_A = 1.8$ V

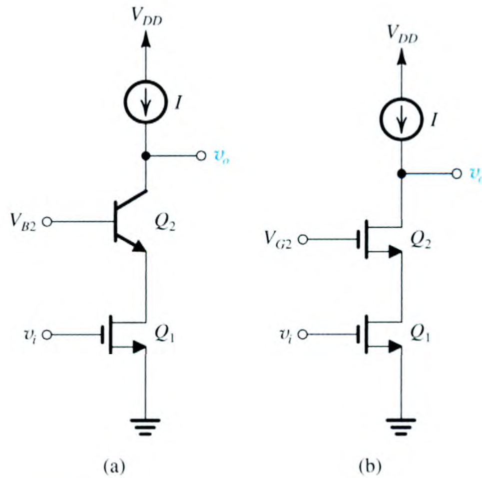


Figure P8.79

Section 8.6: The Integrated-Circuit Source Follower

8.80 A source follower for which $k_n' = 400\ \mu A/V^2$, $V_A' = 9\ V/\mu m$, $\chi = 0.2$, $L = 0.6\ \mu m$, $W = 12\ \mu m$, and $V_t = 0.5$ V is required to provide a dc level shift (between input and output) of 0.8 V. What must the bias current be? Find g_m, g_{mb}, r_o, A_{vo}, and R_o. Assume that the bias current source has an output resistance equal to r_o. Also find the voltage gain when a load resistance of 3 kΩ is connected to the output.

Section 8.7: Current-Mirror Circuits with Improved Performance

8.81 Consider a cascode NMOS mirror using devices fabricated in the 0.13-μm CMOS process specified in Table K.1 (Appendix K). Let all transistors have $W/L = 3.9\ \mu m/0.39\ \mu m$, and let $I_{REF} = 50\ \mu A$. Find the minimum voltage required at the output and the output resistance.

SIM **8.82** A particular cascoded NMOS current mirror, such as that shown in Fig. 8.36, is fabricated in the 0.18-μm CMOS process specified in Table K.1 (Appendix K). All transistors have equal channel lengths $L = 0.54\mu m$. Width $W_1 = W_4 = 2.7\ \mu m$, and $W_2 = W_3 = 27\ \mu m$. The reference current I_{REF} is 20 μA. What output current results? What are the voltages at the gates of Q_2 and Q_3? What is the lowest voltage at the output for which proper current-source operation is possible? What are the values of g_m and r_o of Q_2 and Q_3? What is the output resistance of the mirror?

8.83 To further increase the output resistance of a cascode current source or current mirror, another level of cascoding can be employed as shown is Fig. P8.83. Find the output resistance of this double-cascode current mirror.

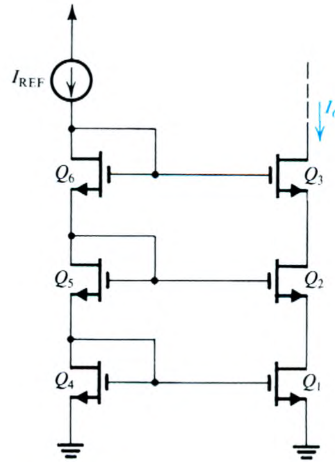


Figure P8.83

8.84 Consider the Wilson current-mirror circuit of Fig. 8.37(a) when supplied with a reference current I_{REF} of 0.2 mA. Find the output resistance R_o. What is the change in I_O corresponding to a change of +1 V in the voltage at the collector of Q_3? Give both the absolute value and the percentage change. Let $\beta = 100$ and $V_A = 20$ V.

D 8.85 (a) The circuit in Fig. P8.85 is a modified version of the Wilson current mirror. Here the output transistor is "split" into two matched transistors, Q_3 and Q_4. Find I_{O1} and I_{O2} in terms of I_{REF}. Assume all transistors to be matched with current gain β.

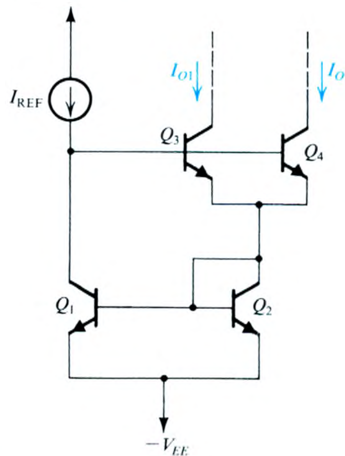


Figure P8.85

(b) Use this idea to design a circuit that generates currents of 0.1 mA, 0.2 mA, and 0.4 mA, using a reference current source of 0.7 mA. What are the actual values of the currents generated for $\beta = 50$?

D 8.86 Use the *pnp* version of the Wilson current mirror to design a 0.1-mA current source. The current source is

required to operate with the voltage at its output terminal as low as –1.5 V. If the power supplies available are ±1.5 V, what is the highest voltage possible at the output terminal?

***8.87** For the Wilson current mirror of Fig. 8.37(a), show that the incremental input resistance seen by I_{REF} is approximately $2V_T/I_{REF}$. (Neglect the Early effect in this derivation and assume a signal ground at the output.) Evaluate R_{in} for $I_{REF} = 0.2$ mA.

8.88 The Wilson MOS mirror in Fig. 8.38(a) is fabricated in the 0.13-μm CMOS process specified in Table K.1 (Appendix K). All three transistors are identical with channel length $L = 4L_{min}$ and $W/L = 16$. For $I_{REF} = 100$ μA, find the voltage at the gates of Q_1 and Q_2, and the voltage at the gate of Q_3. What is the minimum voltage required at the drain of Q_3 for proper operation of the mirror? Find the output resistance R_o.

8.89 Show that the incremental input resistance (seen by I_{REF}) for the Wilson MOS mirror of Fig. 8.38(a) is $2/g_m$. Assume that all three transistors are identical and neglect the Early effect. Also, assume a signal ground at the output. (*Hint*: Replace all transistors by their T model and remember that Q_1 is equivalent to a resistance $1/g_m$.)

***8.90** Consider the Wilson MOS mirror of Fig. 8.88(a) for the case of all transistors identical, with $W/L = 10$, $\mu_n C_{ox} = 400\ \mu A/V^2$, $V_{tn} = 0.5$ V, and $V_A = 18$ V. The mirror is fed with $I_{REF} = 180\ \mu$A.

(a) Obtain an estimate of V_{OV} and V_{GS} at which the three transistors are operating, by neglecting the Early effect.

(b) Noting that Q_1 and Q_2 are operating at different V_{DS}, obtain an approximate value for the difference in their currents and hence determine I_O.

(c) To eliminate the systematic error between I_O and I_{REF} caused by the difference in V_{DS} between Q_1 and Q_2, a diode-connected transistor Q_4 can be added to the circuit as shown in Fig. 8.38(c). What do you estimate I_O to be now?

(d) What is the minimum allowable voltage at the output node of the mirror?

(e) Convince yourself that Q_4 will have no effect on the output resistance of the mirror. Find R_o.

(f) What is the change in I_O (both absolute value and percentage) that results from $\Delta V_O = 1$ V?

8.91 Using the idea employed in the modified Wilson MOS mirror in Fig. 8.38(c), modify the Wilson BJT mirror of Fig. 8.37(a) to eliminate the systematic error resulting from the difference in V_{CE} of Q_1 and Q_2.

D 8.92 (a) Utilizing a reference current of 200 μA, design a Widlar current source to provide an output current of 20 μA. Assume β to be high.

(b) If $\beta = 200$ and $V_A = 50$ V, find the value of the output resistance, and find the change in output current corresponding to a 5-V change in output voltage.

D 8.93 (a) For the circuit in Fig. P8.93, assume BJTs with high β and $v_{BE} = 0.7$ V at 1 mA. Find the value of R that will result in $I_O = 10\ \mu$A.

(b) For the design in (a), find R_o assuming $\beta = 100$ and $V_A = 50$ V.

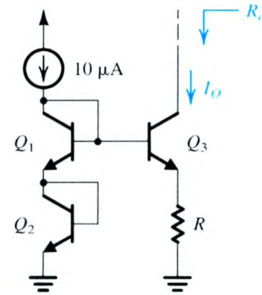


Figure P8.93

D 8.94 If the *pnp* transistor in the circuit of Fig. P8.94 is characterized by its exponential relationship with a scale current I_S, show that the dc current I is determined by $IR = V_T \ln(I/I_S)$. Assume Q_1 and Q_2 to be matched and Q_3, Q_4, and Q_5 to be matched. Find the value of R that yields a current $I = 100\ \mu$A. For the BJT, $V_{EB} = 0.7$ V at $I_E = 1$ mA.

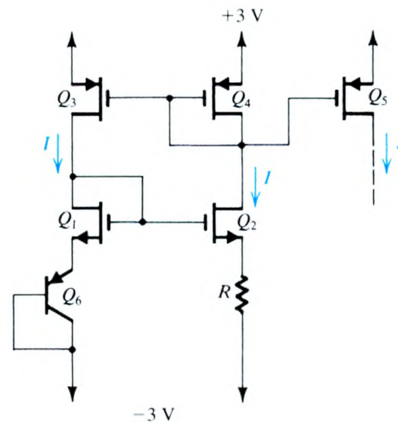


Figure P8.94

CHAPTER 9

Differential and Multistage Amplifiers

IN THIS CHAPTER YOU WILL LEARN

- The essence of the operation of the MOS and the bipolar differential amplifiers: how they reject common-mode noise or interference and amplify differential signals.
- The analysis and design of MOS and BJT differential amplifiers.
- Differential-amplifier circuits of varying complexity, using passive resistive loads, current-source loads, and cascodes—the building blocks we studied in Chapter 8.
- An ingenious and highly popular differential-amplifier circuit that utilizes a current-mirror load.
- The structure, analysis, and design of amplifiers composed of two or more stages in cascade. Two practical examples are studied in detail: a two-stage CMOS op amp and a four-stage bipolar op amp.

Introduction

The differential-pair or differential-amplifier configuration is the most widely used building block in analog integrated-circuit design. For instance, the input stage of every op amp is a differential amplifier. Also, the BJT differential amplifier is the basis of a very-high-speed logic-circuit family, called emitter-coupled logic (ECL).

Initially invented in the 1940s for use with vacuum tubes, the basic differential-amplifier configuration was subsequently implemented with discrete bipolar transistors. However, it was the advent of integrated circuits that has made the differential pair extremely popular in both bipolar and MOS technologies. There are two reasons why differential amplifiers are so well suited for IC fabrication: First, as we shall see, the performance of the differential pair depends critically on the matching between the two sides of the circuit. Integrated-circuit fabrication is capable of providing matched devices whose parameters track over wide ranges of changes in environmental conditions. Second, by their very nature, differential amplifiers utilize more components (approaching twice as many) than single-ended circuits. Integrated-circuit technology makes available large numbers of transistors at relatively low cost.

We assume that the reader is familiar with the basic concept of a differential amplifier as presented in Section 2.1. Nevertheless it is worthwhile to answer the question: Why differential? Basically, there are two reasons for using differential in preference to single-ended amplifiers. First, differential circuits are much less sensitive to noise and interference than single-ended circuits. Consider, for instance, two wires carrying a small differential signal as the voltage difference between the two wires. Now, assume that there is an interference signal that is coupled to the two wires, either capacitively or inductively. As the two wires are physically close together, the interference voltages on the two wires (i.e., between each of the two wires and ground) will be equal. Since, in a differential system, only the difference signal between the two wires is sensed, it will contain no interference component!

The second reason for preferring differential amplifiers is that the differential configuration enables us to bias the amplifier and to couple amplifier stages together without the need for bypass and coupling capacitors such as those utilized in the design of discrete-circuit amplifiers (Section 7.5). This is another reason why differential circuits are ideally suited for IC fabrication where large capacitors are impossible to fabricate economically.

The major topic of this chapter is the differential amplifier in both its MOS and bipolar implementations. You will see that the design and analysis of differential amplifiers relies heavily on the material on single-stage amplifiers presented in Chapters 7 and 8. We will follow the study of differential amplifiers with examples of practical multistage amplifiers, again in both MOS and bipolar technologies.

9.1 The MOS Differential Pair

Figure 9.1 shows the basic MOS differential-pair configuration. It consists of two matched transistors, Q_1 and Q_2, whose sources are joined together and biased by a constant-current source I. The latter is usually implemented by a MOSFET circuit of the type studied in Sections 8.2 and 8.5. For the time being, let's assume that the current source is ideal with infinite output resistance. Although each drain is shown connected to the positive supply through a resistance R_D, in most cases active (current-source) loads are used, as we will see

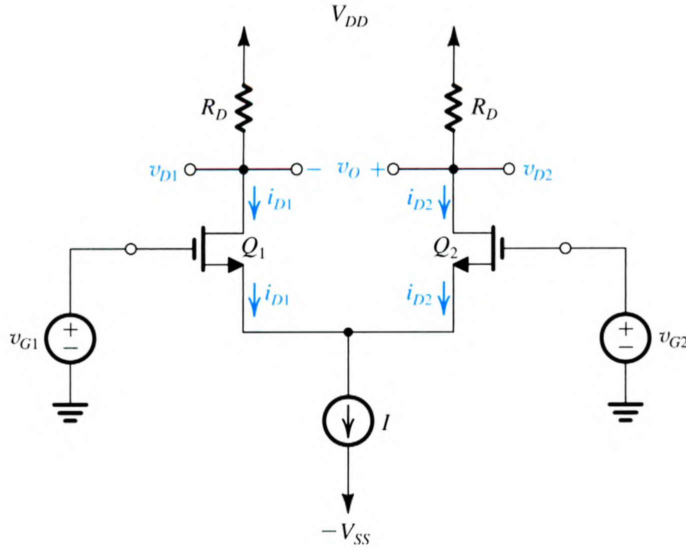


Figure 9.1 The basic MOS differential-pair configuration.

shortly. For the time being, however, we will explain the essence of the differential-pair operation using simple resistive loads. Whatever type of load is used, it is essential that the MOSFETs operate in the saturation region at all times.

The differential amplifier in Fig. 9.1 is fed with two voltages, v_{G1} and v_{G2}, which can in general be expressed as

$$v_{G1} = V_{CM} + \left(\frac{v_{id}}{2}\right)$$
$$v_{G2} = V_{CM} - \left(\frac{v_{id}}{2}\right)$$

where V_{CM}, the average of v_{G1} and v_{G2}, is the **common-mode** input, and v_{id}, the difference between v_{G1} and v_{G2}, is the **differential** input.

9.1.1 Operation with a Common-Mode Input Voltage

To see how the differential pair works, consider first the case when equal dc voltages, V_{CM}, are applied to the two gate terminals. That is, as shown in Fig. 9.2, $v_{G1} = v_{G2} = V_{CM}$. Since Q_1 and Q_2 are matched, the current I will divide equally between the two transistors. Thus, $i_{D1} = i_{D2} = I/2$, and the voltage at the sources, V_S, will be

$$V_S = V_{CM} - V_{GS} \tag{9.1}$$

where V_{GS} is the gate-to-source voltage corresponding to a drain current of $I/2$. Neglecting channel-length modulation, V_{GS} and $I/2$ are related by

$$\frac{I}{2} = \frac{1}{2}k_n' \frac{W}{L}(V_{GS} - V_t)^2 \tag{9.2}$$

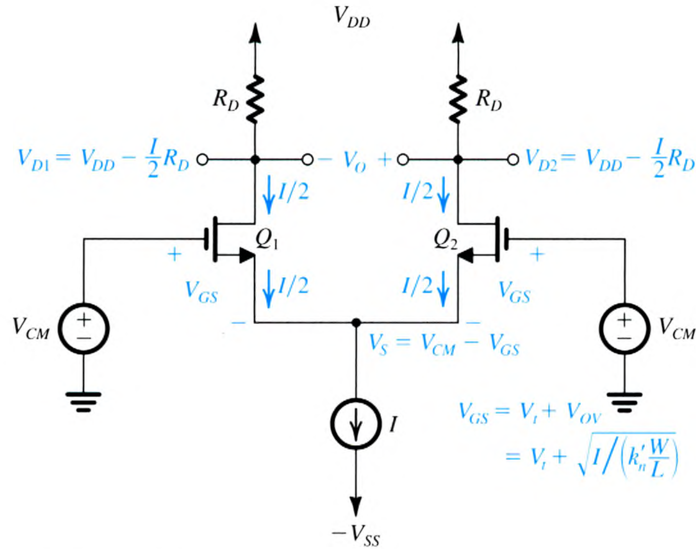


Figure 9.2 The MOS differential pair with a common-mode input voltage V_{CM}.

or in terms of the overdrive voltage V_{OV},

$$V_{OV} = V_{GS} - V_t \tag{9.3}$$

$$\frac{I}{2} = \frac{1}{2}k_n'\frac{W}{L}V_{OV}^2 \tag{9.4}$$

$$\Rightarrow \qquad V_{OV} = \sqrt{I/k_n'(W/L)} \tag{9.5}$$

The voltage at each drain will be

$$V_{D1} = V_{D2} = V_{DD} - \frac{I}{2}R_D \tag{9.6}$$

and the output voltage between the two drains

$$V_O = V_{D2} - V_{D1}$$

will be zero.

Now, let us vary the value of the common-mode voltage V_{CM}. We see that the current I will continue to divide equally between Q_1 and Q_2, and thus the voltages at the drains will not change and the output voltage V_O will continue to be zero. Thus the differential pair does *not* respond to (i.e., it *rejects*) common-mode input signals.

An important specification of a differential amplifier is its **input common-mode range**. This is the range of V_{CM} over which the differential pair operates *properly*. The highest value of V_{CM} is limited by the requirement that Q_1 and Q_2 remain in saturation, thus

$$V_{CM\max} = V_t + V_{D1,2} = V_t + V_{DD} - \frac{I}{2}R_D \tag{9.7}$$

The lowest value of V_{CM} is determined by the need to allow for a sufficient voltage across the current source I for it to operate properly. If a voltage V_{CS} is needed across the current source, then

$$V_{CM\min} = -V_{SS} + V_{CS} + V_t + V_{OV} \tag{9.8}$$

Example 9.1

For the MOS differential pair with a common-mode voltage V_{CM} applied, as shown in Fig. 9.2, let $V_{DD} = V_{SS} = 1.5$ V, $k_n'(W/L) = 4$ mA/V^2, $V_t = 0.5$ V, $I = 0.4$ mA, and $R_D = 2.5$ kΩ, and neglect channel-length modulation. Assume that the current source I requires a minimum voltage of 0.4 V to operate properly.

(a) Find V_{OV} and V_{GS} for each transistor.
(b) For $V_{CM} = 0$, find V_S, I_{D1}, I_{D2}, V_{D1}, V_{D2}, and V_O.
(c) Repeat (b) for $V_{CM} = +1$ V.
(d) Repeat (b) for $V_{CM} = -0.2$ V.
(e) What is the highest permitted value of V_{CM}?
(f) What is the lowest value allowed for V_{CM}?

Solution

(a) With $v_{G1} = v_{G2} = V_{CM}$, we see that $V_{GS1} = V_{GS2}$. Now, since the transistors are matched, I will divide equally between the two transistors,

$$I_{D1} = I_{D2} = \frac{I}{2}$$

Thus,

$$\frac{I}{2} = \frac{1}{2}k_n'(W/L)V_{OV}^2$$

$$\frac{0.4}{2} = \frac{1}{2} \times 4V_{OV}^2$$

which results in

$$V_{OV} = 0.316 \text{ V}$$

and thus,

$$V_{GS} = V_t + V_{OV} = 0.5 + 0.316 \simeq 0.82 \text{ V}$$

(b) The analysis for the case $V_{CM} = 0$ is shown in Fig. 9.3(a), from which we see that

$$V_S = V_G - V_{GS} = 0 - 0.82 = -0.82 \text{ V}$$

$$I_{D1} = I_{D2} = \frac{I}{2} = 0.2 \text{ mA}$$

$$V_{D1} = V_{D2} = V_{DD} - \frac{I}{2}R_D$$

$$= 1.5 - 0.2 \times 2.5 = 1 \text{ V}$$

$$V_O = V_{D2} - V_{D1} = 0 \text{ V}$$

(c) The analysis for the case $V_{CM} = +1$ V is shown in Fig. 9.3(b), from which we see that

$$V_S = V_G - V_{GS} = 1 - 0.82 = +0.18 \text{ V}$$

$$I_{D1} = I_{D2} = \frac{I}{2} = 0.2 \text{ mA}$$

Example 9.1 *continued*

$$V_{D1} = V_{D2} = V_{DD} - \frac{I}{2} R_D = 1.5 - 0.2 \times 2.5 = +1 \text{ V}$$
$$V_O = V_{D2} - V_{D1} = 0 \text{ V}$$

Observe that the transistors remain in the saturation region as assumed. Also observe that I_{D1}, I_{D2}, V_{D1}, and V_{D2}, and V_O remain unchanged even though the common-mode voltage V_{CM} changed by 1 V.

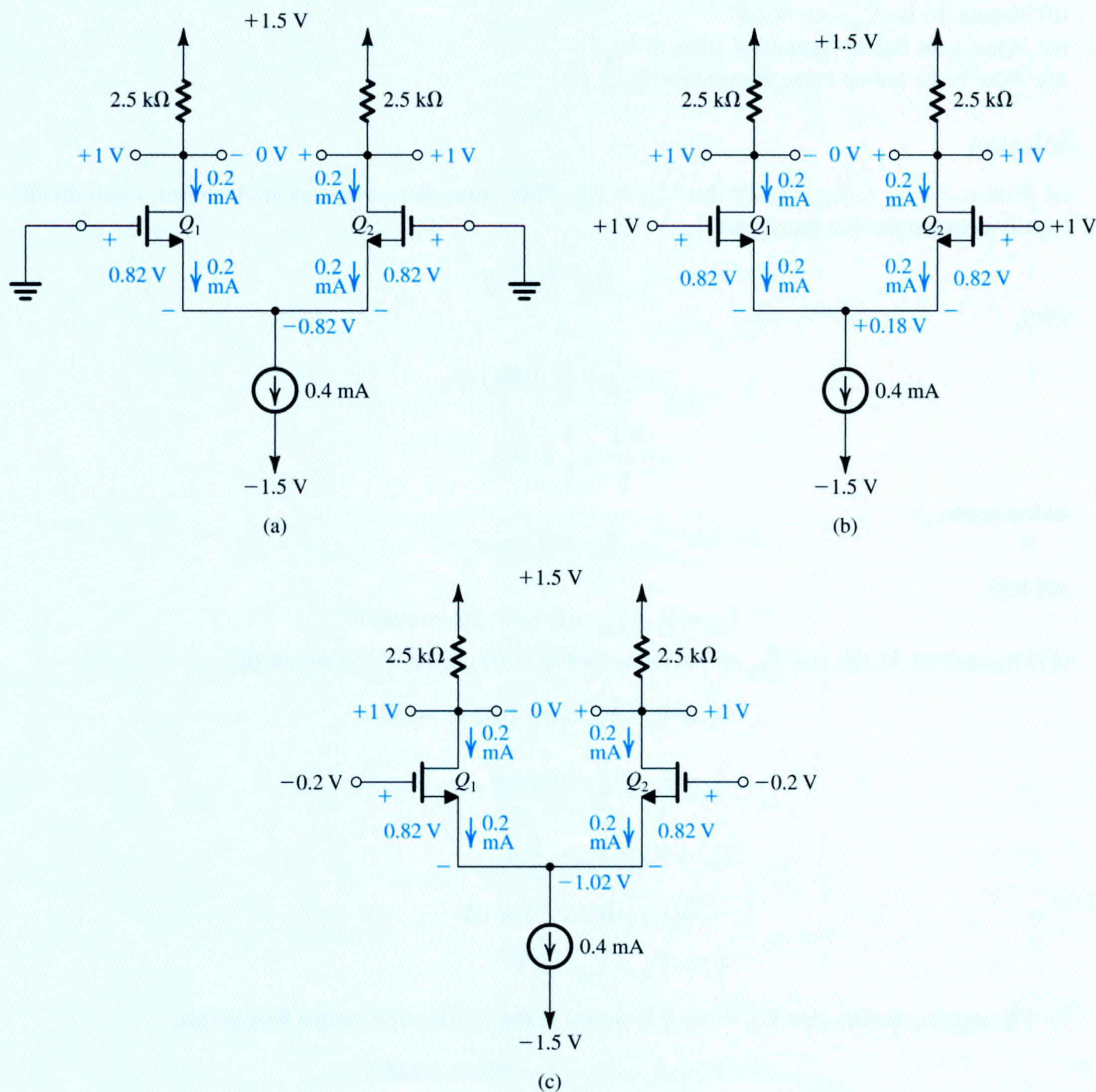


Figure 9.3 Circuits for Example 9.1. Effects of varying V_{CM} on the operation of the differential pair.

(d) The analysis for the case $V_{CM} = -0.2$ V is shown in Fig. 9.3(c), from which we see that

$$V_S = V_G - V_{GS} = -0.2 - 0.82 = -1.02 \text{ V}$$

It follows that the current source I now has a voltage across it of

$$V_{CS} = -V_S - (-V_{SS}) = -1.02 + 1.5 = 0.48 \text{ V}$$

which is greater than the minimum required value of 0.4 V. Thus, the current source is still operating properly and delivering a constant current $I = 0.4$ mA, and hence

$$I_{D1} = I_{D2} = \frac{I}{2} = 0.2 \text{ mA}$$

$$V_{D1} = V_{D2} = V_{DD} - \frac{I}{2} R_D = +1 \text{ V}$$

$$V_O = V_{D2} - V_{D1} = 0 \text{ V}$$

So, here again the differential circuit is not responsive to the change in the common-mode voltage V_{CM}.

(e) The highest permitted value of V_{CM} is that which causes Q_1 and Q_2 to leave saturation and enter the triode region. Thus,

$$V_{CM\max} = V_t + V_D$$
$$= 0.5 + 1 = +1.5 \text{ V}$$

(f) The lowest value allowed for V_{CM} is that which reduces the voltage across the current source I to the minimum required of $V_{CS} = 0.4$ V. Thus,

$$V_{CM\min} = -V_{SS} + V_{CS} + V_{GS}$$
$$= -1.5 + 0.4 + 0.82 = -0.28 \text{ V}$$

Thus, the input common-mode range is

$$-0.28 \text{ V} \le V_{CM} \le +1.5 \text{ V}$$

EXERCISE

9.1 For the amplifier in Example 9.1, find the input common-mode range for the case in which the two drain resistances R_D are increased by a factor of 2.
Ans. −0.28 V to 1.0 V

9.1.2 Operation with a Differential Input Voltage

Next we apply a difference or differential input voltage by grounding the gate of Q_2 (i.e., setting $v_{G2} = 0$) and applying a signal v_{id} to the gate of Q_1, as shown in Fig. 9.4. We can see that since $v_{id} = v_{GS1} - v_{GS2}$, if v_{id} is positive, v_{GS1} will be greater than v_{GS2} and hence i_{D1} will be greater than i_{D2}; correspondingly, v_{D1} will be lower than v_{D2}, and the difference output voltage v_O will be positive. On the other hand, when v_{id} is negative, v_{GS1} will be lower than v_{GS2}, i_{D1} will be smaller than i_{D2}, and correspondingly v_{D1} will be higher than v_{D2}; in other words, the difference or differential output voltage v_O will be negative.

From the above, we see that the differential pair responds to **difference-mode** or **differential input signals** by providing a corresponding differential output signal between the two drains. At this point, it is useful to inquire about the value of v_{id} that causes the entire bias current I to flow in one of the two transistors. In the positive direction, this happens when v_{GS1} reaches the value that corresponds to $i_{D1} = I$, and v_{GS2} is reduced to a value equal to the threshold voltage V_t, at which point $v_S = -V_t$. The value of v_{GS1} can be found from

$$I = \frac{1}{2}\left(k_n' \frac{W}{L}\right)(v_{GS1} - V_t)^2$$

as

$$\begin{aligned} v_{GS1} &= V_t + \sqrt{2I/k_n'(W/L)} \\ &= V_t + \sqrt{2}V_{OV} \end{aligned} \tag{9.9}$$

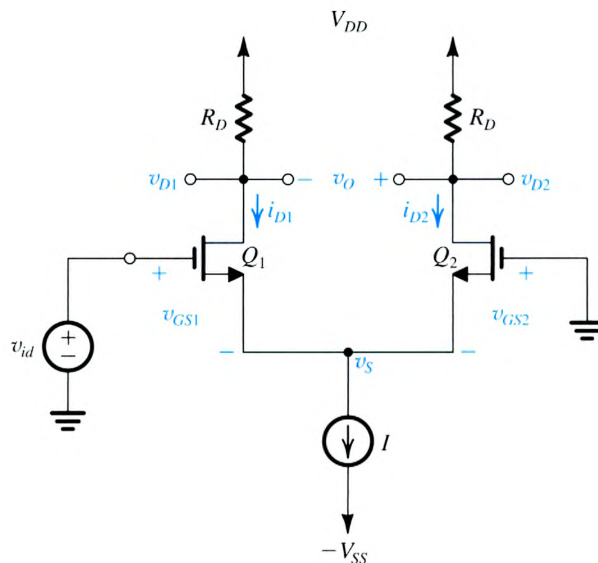


Figure 9.4 The MOS differential pair with a differential input signal v_{id} applied. With v_{id} positive: $v_{GS1} > v_{GS2}$, $i_{D1} > i_{D2}$, and $v_{D1} < v_{D2}$; thus v_O will be positive. With v_{id} negative: $v_{GS1} < v_{GS2}$, $i_{D1} < i_{D2}$, and $v_{D1} > v_{D2}$; thus v_O will be negative.

where V_{OV} is the overdrive voltage corresponding to a drain current of $I/2$ (Eq. 9.5). Thus, the value of v_{id} at which the entire bias current I is steered into Q_1 is

$$\begin{aligned} v_{id\max} &= v_{GS1} + v_S \\ &= V_t + \sqrt{2}V_{OV} - V_t \\ &= \sqrt{2}V_{OV} \end{aligned} \tag{9.10}$$

If v_{id} is increased beyond $\sqrt{2}V_{OV}$, i_{D1} remains equal to I, v_{GS1} remains equal to $(V_t + \sqrt{2}V_{OV})$, and v_S rises correspondingly, thus keeping Q_2 off. In a similar manner we can show that in the negative direction, as v_{id} reaches $-\sqrt{2}V_{OV}$, Q_1 turns off and Q_2 conducts the entire bias current I. Thus the current I can be steered from one transistor to the other by varying v_{id} in the range

$$-\sqrt{2}V_{OV} \leq v_{id} \leq \sqrt{2}V_{OV}$$

which defines the range of differential-mode operation. Finally, observe that we have assumed that Q_1 and Q_2 remain in saturation even when one of them is conducting the entire current I.

EXERCISE

9.2 For the MOS differential pair specified in Example 9.1 find (a) the value of v_{id} that causes Q_1 to conduct the entire current I, and the corresponding values of v_{D1}, v_{D2}, and v_O; (b) the value of v_{id} that causes Q_2 to conduct the entire current I, and the corresponding values of v_{D1}, v_{D2}, and v_O; (c) the corresponding range of the differential output voltage v_O.
Ans. (a) +0.45 V, 0.5 V, 1.5 V, +1 V; (b) –0.45 V, 1.5 V, 0.5 V, −1 V; (c) +1 V to –1 V

To use the differential pair as a linear amplifier, we keep the differential input signal v_{id} small. As a result, the current in one of the transistors (Q_1 when v_{id} is positive) will increase by an increment ΔI proportional to v_{id}, to $(I/2 + \Delta I)$. Simultaneously, the current in the other transistor will decrease by the same amount to become $(I/2 - \Delta I)$. A voltage signal $-\Delta I R_D$ develops at one of the drains and an opposite-polarity signal, $\Delta I R_D$, develops at the other drain. Thus the output voltage taken between the two drains, v_O, will be $2\Delta I\, R_D$, which is proportional to the differential input signal v_{id}. The small-signal operation of the differential pair will be studied in detail in Section 9.1.4.

9.1.3 Large-Signal Operation

We shall now derive expressions for the drain currents i_{D1} and i_{D2} in terms of the input differential signal $v_{id} \equiv v_{G1} - v_{G2}$. The derivation assumes that the differential pair is perfectly matched and neglects channel-length modulation ($\lambda = 0$ or $V_A = \infty$). Thus these expressions do not depend on the details of the circuit to which the drains are connected, and we do not show these connections in Fig. 9.5; we simply assume that the circuit maintains Q_1 and Q_2 in the saturation region of operation at all times.

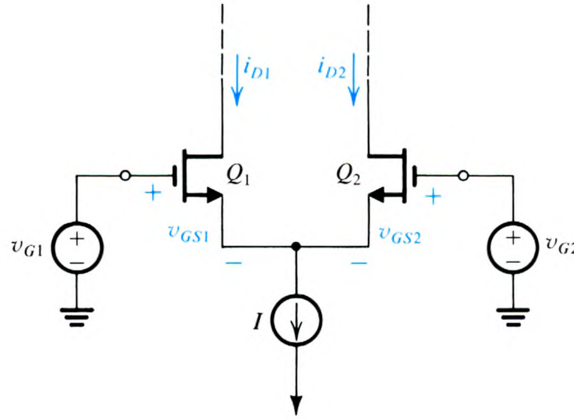


Figure 9.5 The MOSFET differential pair for the purpose of deriving the transfer characteristics, i_{D1} and i_{D2} versus $v_{id} = v_{G1} - v_{G2}$.

To begin with, we express the drain currents of Q_1 and Q_2 as

$$i_{D1} = \frac{1}{2}k_n'\frac{W}{L}(v_{GS1} - V_t)^2 \tag{9.11}$$

$$i_{D2} = \frac{1}{2}k_n'\frac{W}{L}(v_{GS2} - V_t)^2 \tag{9.12}$$

Taking the square roots of both sides of each of Eqs. (9.11) and (9.12), we obtain

$$\sqrt{i_{D1}} = \sqrt{\frac{1}{2}k_n'\frac{W}{L}}(v_{GS1} - V_t) \tag{9.13}$$

$$\sqrt{i_{D2}} = \sqrt{\frac{1}{2}k_n'\frac{W}{L}}(v_{GS2} - V_t) \tag{9.14}$$

Subtracting Eq. (9.14) from Eq. (9.13) and substituting

$$v_{GS1} - v_{GS2} = v_{G1} - v_{G2} = v_{id} \tag{9.15}$$

results in

$$\sqrt{i_{D1}} - \sqrt{i_{D2}} = \sqrt{\frac{1}{2}k_n'\frac{W}{L}}\,v_{id} \tag{9.16}$$

The constant-current bias imposes the constraint

$$i_{D1} + i_{D2} = I \tag{9.17}$$

Equations (9.16) and (9.17) are two equations in the two unknowns i_{D1} and i_{D2} and can be solved as follows: Squaring both sides of Eq. (9.16) and substituting for $i_{D1} + i_{D2} = I$ gives

$$2\sqrt{i_{D1}i_{D2}} = I - \frac{1}{2}k_n'\frac{W}{L}v_{id}^2$$

Substituting for i_{D2} from Eq. (9.17) as $i_{D2} = I - i_{D1}$ and squaring both sides of the resulting equation provides a quadratic equation in i_{D1} that can be solved to yield

$$i_{D1} = \frac{I}{2} \pm \sqrt{k_n'\frac{W}{L}I}\left(\frac{v_{id}}{2}\right)\sqrt{1 - \frac{(v_{id}/2)^2}{I/k_n'\frac{W}{L}}}$$

Now, since the increment in i_{D1} above the bias value of $I/2$ must have the same polarity as v_{id}, only the root with the "+" sign in the second term is physically meaningful; thus,

$$i_{D1} = \frac{I}{2} + \sqrt{k_n' \frac{W}{L} I}\left(\frac{v_{id}}{2}\right)\sqrt{1 - \frac{(v_{id}/2)^2}{I/k_n' \frac{W}{L}}} \tag{9.18}$$

The corresponding value of i_{D2} is found from $i_{D2} = I - i_{D1}$ as

$$i_{D2} = \frac{I}{2} - \sqrt{k_n' \frac{W}{L} I}\left(\frac{v_{id}}{2}\right)\sqrt{1 - \frac{(v_{id}/2)^2}{I/k_n' \frac{W}{L}}} \tag{9.19}$$

At the bias (quiescent) point, $v_{id} = 0$, leading to

$$i_{D1} = i_{D2} = \frac{I}{2} \tag{9.20}$$

Correspondingly,

$$v_{GS1} = v_{GS2} = V_{GS} \tag{9.21}$$

where

$$\frac{I}{2} = \frac{1}{2} k_n' \frac{W}{L}(V_{GS} - V_t)^2 = \frac{1}{2} k_n' \frac{W}{L} V_{OV}^2 \tag{9.22}$$

This relationship enables us to replace $k_n'(W/L)$ in Eqs. (9.18) and (9.19) with I/V_{OV}^2 to express i_{D1} and i_{D2} in the alternative form

$$i_{D1} = \frac{I}{2} + \left(\frac{I}{V_{OV}}\right)\left(\frac{v_{id}}{2}\right)\sqrt{1 - \left(\frac{v_{id}/2}{V_{OV}}\right)^2} \tag{9.23}$$

$$i_{D2} = \frac{I}{2} - \left(\frac{I}{V_{OV}}\right)\left(\frac{v_{id}}{2}\right)\sqrt{1 - \left(\frac{v_{id}/2}{V_{OV}}\right)^2} \tag{9.24}$$

These two equations describe the effect of applying a differential input signal v_{id} on the currents i_{D1} and i_{D2}. They can be used to obtain the normalized plots, i_{D1}/I and i_{D2}/I versus v_{id}/V_{OV}, shown in Fig. 9.6. Note that at $v_{id} = 0$, the two currents are equal to $I/2$. Making v_{id} positive causes i_{D1} to increase and i_{D2} to decrease by equal amounts, to keep the sum constant, $i_{D1} + i_{D2} = I$. The current is steered entirely into Q_1 when v_{id} reaches the value $\sqrt{2}V_{OV}$, as we found out earlier. For v_{id} negative, identical statements can be made by interchanging i_{D1} and i_{D2}. In this case, $v_{id} = -\sqrt{2}V_{OV}$ steers the current entirely into Q_2. Finally, note that the plots in Fig. 9.6 are universal, as they apply to any MOS differential pair.

The transfer characteristics of Eqs. (9.23) and (9.24) and Fig. 9.6 are obviously nonlinear. This is due to the term involving v_{id}^2. Since we are interested in obtaining linear amplification from the differential pair, we will strive to make this term as small as possible. For a given value of V_{OV}, the only thing we can do is keep $(v_{id}/2)$ much smaller than V_{OV},

$$v_{id} \ll 2V_{OV}$$

which is the condition for the small-signal approximation. It results in

$$i_{D1} \simeq \frac{I}{2} + \left(\frac{I}{V_{OV}}\right)\left(\frac{v_{id}}{2}\right) \tag{9.25}$$

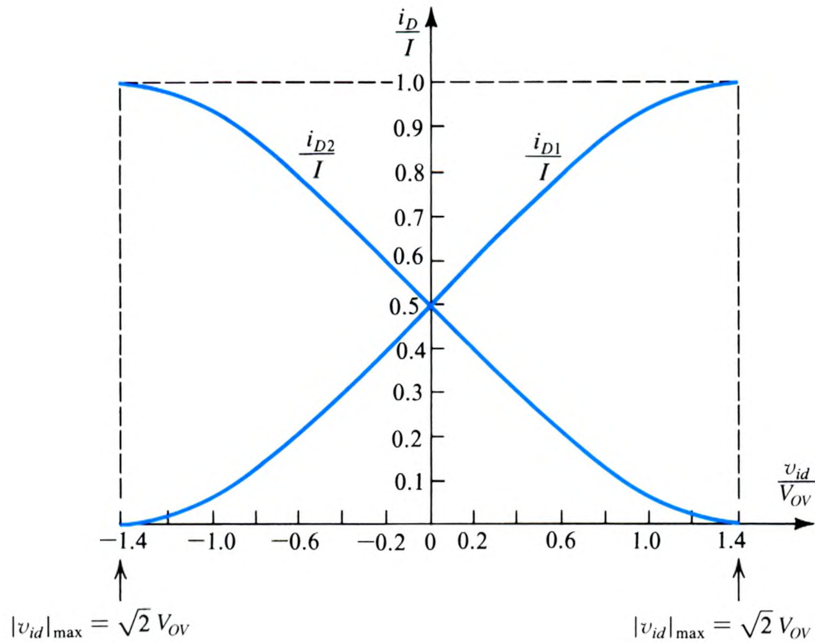


Figure 9.6 Normalized plots of the currents in a MOSFET differential pair. Note that V_{OV} is the overdrive voltage at which Q_1 and Q_2 operate when conducting drain currents equal to $I/2$, the equilibrium situation. These graphs are universal and apply to any MOS differential pair.

and

$$i_{D2} \simeq \frac{I}{2} - \left(\frac{I}{V_{OV}}\right)\left(\frac{v_{id}}{2}\right) \tag{9.26}$$

which, as expected, indicate that i_{D1} increases by an increment i_d, and i_{D2} decreases by the same amount, i_d, where i_d is proportional to the differential input signal v_{id},

$$i_d = \left(\frac{I}{V_{OV}}\right)\left(\frac{v_{id}}{2}\right) \tag{9.27}$$

Recalling from our study of the MOSFET amplifier in Chapter 7 that a MOSFET biased at a current I_D has a transconductance $g_m = 2I_D/V_{OV}$, we recognize the factor (I/V_{OV}) in Eq. (9.27) as g_m of each of Q_1 and Q_2, which are biased at $I_D = I/2$. Now, why $v_{id}/2$? Simply because v_{id} divides equally between the two devices with $v_{gs1} = v_{id}/2$ and $v_{gs2} = -v_{id}/2$, which causes Q_1 to have a current increment i_d and Q_2 to have a current decrement i_d. We shall analyze the small-signal operation of the MOS differential pair shortly. At this time, however, we wish to return to Eqs. (9.23) and (9.24) and note that for a given v_{id}, and a given I, linearity can be increased by increasing the overdrive voltage V_{OV} at which each of Q_1 and Q_2 is operating. This can be done by using smaller W/L ratios. The price paid for the increased linearity is a reduction in g_m and hence a reduction in gain. In this regard, we observe that the normalized plot of Fig. 9.6, though compact, masks this design degree of freedom. Figure 9.7 shows plots of the transfer characteristics $i_{D1,2}/I$ versus v_{id} for various values of V_{OV}. These graphs clearly illustrate the linearity–transconductance trade-off obtained by changing the

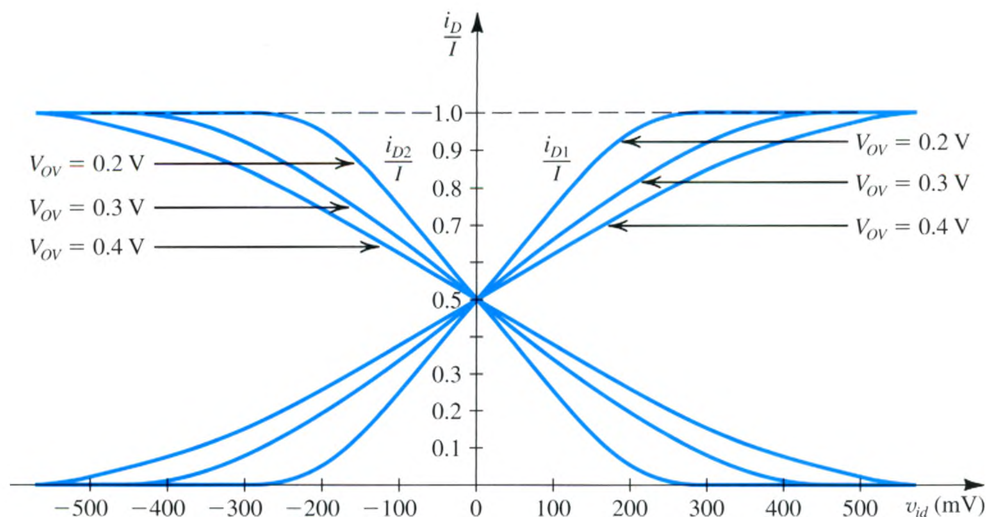


Figure 9.7 The linear range of operation of the MOS differential pair can be extended by operating the transistor at a higher value of V_{OV}.

value of V_{OV}: The linear range of operation can be extended by operating the MOSFETs at a higher V_{OV} (by using smaller W/L ratios) at the expense of reducing g_m and hence the gain. This trade-off is based on the assumption that the bias current I is being kept constant. The bias current can, of course, be increased to obtain a higher g_m. The expense for doing this, however, is increased power dissipation, a serious limitation in IC design.

EXERCISE

9.3 A MOS differential pair is operated at a bias current I of 0.4 mA. If $\mu_n C_{ox} = 0.2\text{ mA/V}^2$, find the required values of W/L and the resulting g_m if the MOSFETs are operated at $V_{OV} = 0.2$, 0.3, and 0.4 V. For each value, give the maximum $|v_{id}|$ for which the term involving v_{id}^2 in Eqs. (9.23) and (9.24), namely, $((v_{id}/2)/V_{OV})^2$, is limited to 0.1.

Ans.

V_{OV} (V)	0.2	0.3	0.4
W/L	50	22.2	12.5
g_m (mA/V)	2	1.33	1
$\lvert v_{id}\rvert_{\max}$ (mV)	126	190	253

9.1.4 Small-Signal Operation

In this section we build on our understanding of the basic operation of the differential pair and consider in some detail how it operates as a linear amplifier.

Differential Gain Figure 9.8(a) shows the MOS differential amplifier with input voltages

$$v_{G1} = V_{CM} + \frac{1}{2}v_{id} \tag{9.28}$$

and

$$v_{G2} = V_{CM} - \frac{1}{2}v_{id} \tag{9.29}$$

Here, V_{CM} denotes a common-mode dc voltage within the input common-mode range of the differential amplifier. It is needed in order to set the dc voltage of the MOSFET gates. Typically V_{CM} is at the middle value of the power supply. Thus, for our case, where two complementary supplies are utilized, V_{CM} is typically 0 V.

The differential input signal v_{id} is applied in a **complementary** (or **balanced**) manner; that is, v_{G1} is increased by $v_{id}/2$ and v_{G2} is decreased by $v_{id}/2$. This would be the case, for instance, if the differential amplifier were fed from the output of another differential-amplifier stage. Sometimes, however, the differential input is applied in a

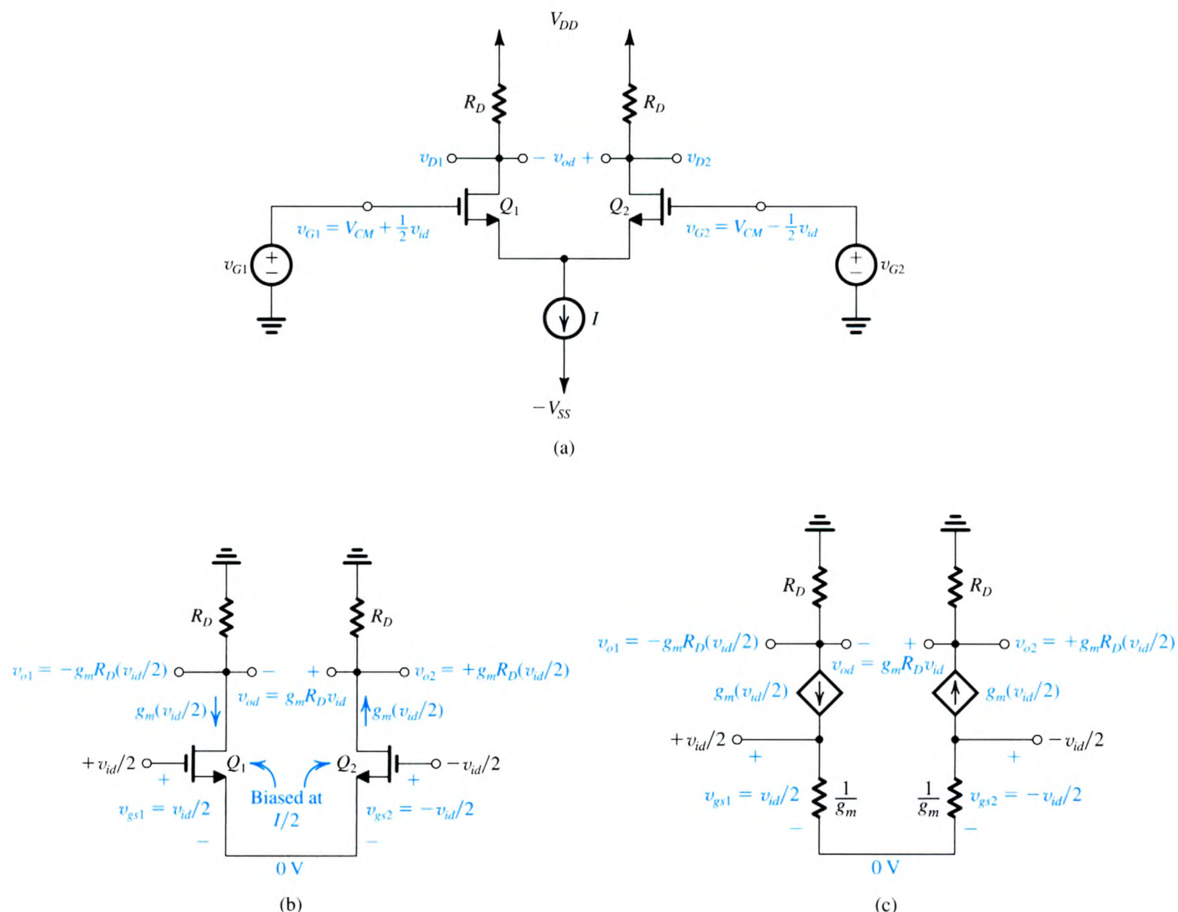


Figure 9.8 Small-signal analysis of the MOS differential amplifier. (**a**) The circuit with a common-mode voltage applied to set the dc bias voltage at the gates and with v_{id} applied in a complementary (or balanced) manner. (**b**) The circuit prepared for small-signal analysis. (**c**) The circuit in (**b**), with the MOSFETs replaced with T models.

single-ended fashion, as we saw earlier in Fig. 9.4. The difference in the resulting performance is too subtle for our current needs.

As Fig. 9.8(a) shows, the amplifier output can be taken either between one of the drains and ground or between the two drains. In the first case, the resulting **single-ended outputs** v_{o1} and v_{o2} will be riding on top of the dc voltages at the drains, $\left(V_{DD} - \frac{I}{2}R_D\right)$. This is not the case when the output is taken between the two drains; the resulting **differential** output v_{od} (having a 0-V dc component) will be entirely a signal component. Later, we will see other significant advantages to taking the output voltage differentially.

Our objective now is to analyze the small-signal operation of the differential amplifier of Fig. 9.8(a) to determine its voltage gain in response to the differential input signal v_{id}. To that end we show in Fig. 9.8(b) the circuit with the power supplies grounded, the bias current source I removed, and V_{CM} eliminated; that is, only signal quantities are shown. For the time being we will neglect the effect of the MOSFET r_o. Finally, note that Q_1 and Q_2 are both biased at a dc current of $I/2$ and operating at an overdrive voltage V_{OV}.

From the symmetry of the circuit and the balanced manner in which v_{id} is applied, we observe that the signal voltage at the joint source connection must be zero, acting as a sort of **virtual ground**. Thus Q_1 has a gate-to-source voltage signal $v_{gs1} = v_{id}/2$, and Q_2 has $v_{gs2} = -v_{id}/2$. Assuming $v_{id}/2 \ll V_{OV}$, the condition for the small-signal approximation, the changes resulting in the drain currents of Q_1, and Q_2 will be proportional to v_{gs1} and v_{gs2}, respectively. Thus Q_1 will have a drain current increment $g_m(v_{id}/2)$, and Q_2 will have a drain current decrement $g_m(v_{id}/2)$, where g_m denotes the equal transconductances of the two devices,

$$g_m = \frac{2I_D}{V_{OV}} = \frac{2(I/2)}{V_{OV}} = \frac{I}{V_{OV}} \tag{9.30}$$

These results correspond to those obtained earlier using the large-signal transfer characteristics and imposing the small-signal condition, Eqs. (9.25) to (9.27). To further illustrate the small-signal operation of the differential amplifier, we show in Fig. 9.8(c) its equivalent circuit obtained by replacing each of the MOSFETs with the corresponding T model. We encourage you to study the correspondence between the elements and quantities in Fig. 9.8(b) and 9.8(c).

It is useful at this point to observe again that a signal ground is established at the source terminals of the transistors *without resorting to the use of a large bypass capacitor*, clearly a major advantage of the differential-pair configuration.

The essence of differential-pair operation is that it provides complementary current signals in the drains; what we do with the resulting pair of complementary current signals is, in a sense, a separate issue. Here, of course, we are simply passing the two current signals through a pair of matched resistors, R_D, and thus obtaining the drain voltage signals

$$v_{o1} = -g_m \frac{v_{id}}{2} R_D \tag{9.31}$$

and

$$v_{o2} = +g_m \frac{v_{id}}{2} R_D \tag{9.32}$$

If the output is taken in a single-ended fashion, the resulting gain becomes

$$\frac{v_{o1}}{v_{id}} = -\frac{1}{2} g_m R_D \tag{9.33}$$

or

$$\frac{v_{o2}}{v_{id}} = \frac{1}{2} g_m R_D \tag{9.34}$$

Alternatively, if the output is taken differentially, the gain becomes

$$A_d \equiv \frac{v_{od}}{v_{id}} = \frac{v_{o2} - v_{o1}}{v_{id}} = g_m R_D \tag{9.35}$$

Thus another advantage of taking the output differentially is an increase in gain by a factor of 2 (6 dB). It should be noted, however, that although differential outputs are preferred, a single-ended output is needed in some applications. We will have more to say about this later.

An alternative and useful way of viewing the operation of the differential pair in response to a differential input signal v_{id} is illustrated in Fig. 9.9. Here we are making use of the fact that the resistance between gate and source of a MOSFET, looking into the source, is $1/g_m$. As a result, between G_1 and G_2 we have a total resistance, in the source circuit, of $2/g_m$. It follows that we can obtain the current i_d simply by dividing v_{id} by $2/g_m$, as indicated in the figure.

The Differential Half-Circuit When a symmetrical differential amplifier is fed with a differential signal in a balanced manner, as in the case in Fig. 9.8, the performance can be determined by considering only half the circuit. The equivalent differential half-circuit is shown in Fig. 9.10. It has a grounded source, a result of the virtual ground that appears on the common sources' terminal of the MOSFETs in the differential pair. Note that Q_1 is operating at a drain bias current of $(I/2)$ and an overdrive voltage V_{OV}.

We can determine the differential gain A_d directly from the half-circuit. For instance, if we wish to take r_o of Q_1 and Q_2 into account, we can use the half-circuit with the following result:

$$A_d = g_m(R_D \| r_o) \tag{9.36}$$

More significantly, we can determine the frequency response of the differential gain by analyzing the half-circuit, as we shall do in Chapter 10.

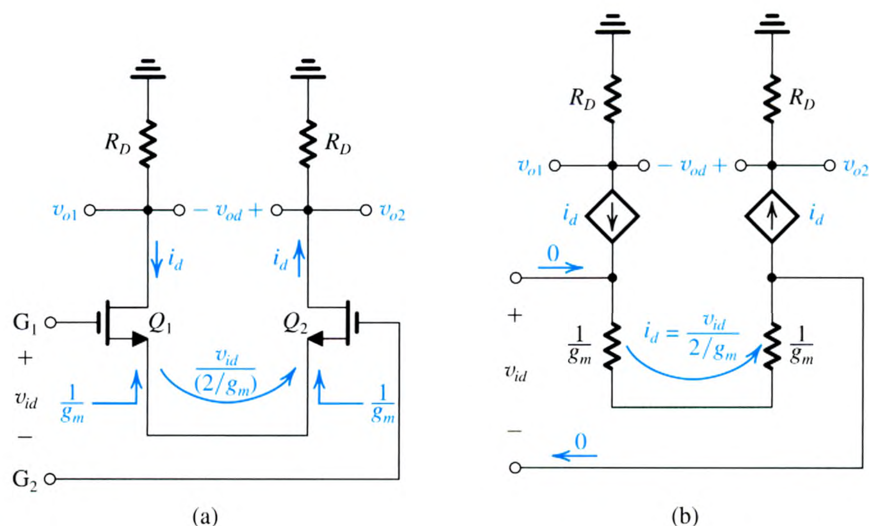


Figure 9.9 An alternative view of the small-signal differential operation of the MOS differential pair: (**a**) analysis done directly on the circuit; (**b**) analysis using equivalent-circuit models.

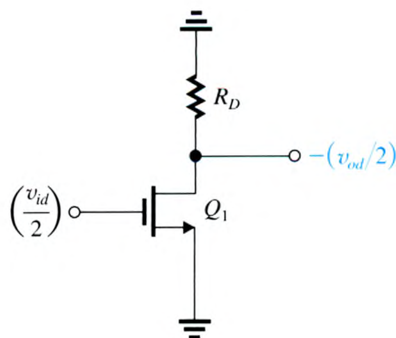


Figure 9.10 The equivalent differential half-circuit of the differential amplifier of Fig. 9.8. Here Q_1 is biased at $I/2$ and is operating at V_{OV}. This circuit can be used to determine the differential voltage gain of the differential amplifier $A_d = v_{od}/v_{id}$.

Example 9.2

Give the differential half-circuit of the differential amplifier shown in Fig. 9.11(a). Assume that Q_1 and Q_2 are perfectly matched. Neglecting r_o, determine the differential voltage gain $A_d \equiv v_{od}/v_{id}$.

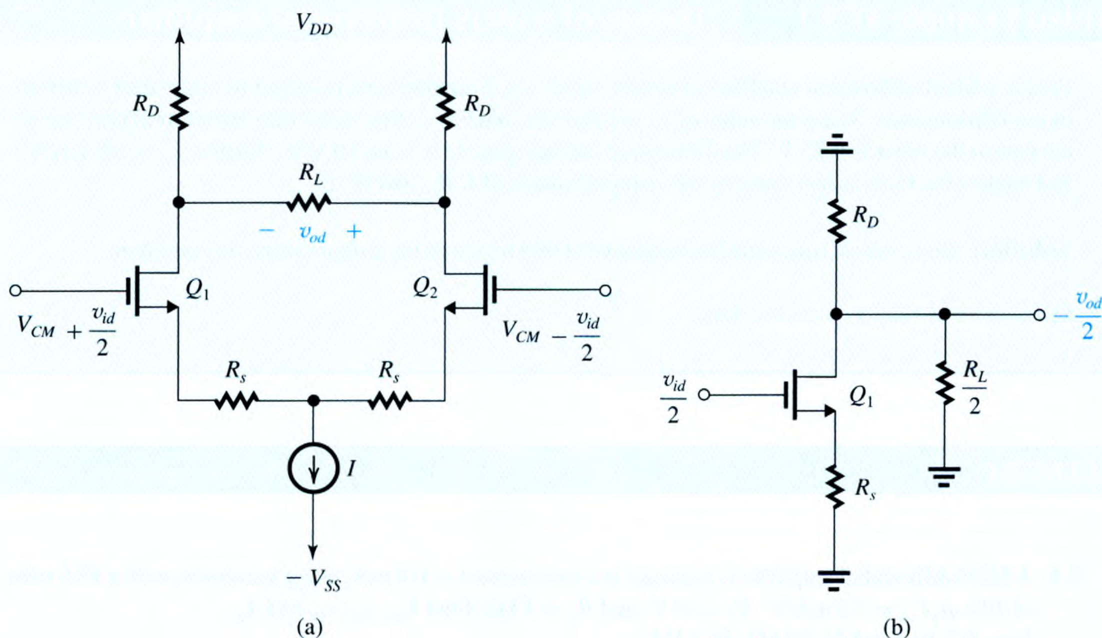


Figure 9.11 (**a**) Differential amplifier for Example 9.2. (**b**) Differential half-circuit.

Solution

Since the circuit is symmetrical and is fed with v_{id} in a balanced manner, the differential half-circuit will be as shown in Fig. 9.11(b). Observe that because the line of symmetry passes through the middle of R_L, the half-circuit has a resistance $R_L/2$ connected between drain and ground. Also note that the virtual ground

Example 9.2 *continued*

appears on the node between the two resistances R_s. As a result, the half-circuit has a source-degeneration resistance R_s.

Now, neglecting r_o of the half-circuit transistor Q_1, we can obtain the gain as the ratio of the total resistance in the drain to the total resistance in the source as

$$\frac{-v_{od}/2}{v_{id}/2} = -\frac{R_D \parallel (R_L/2)}{1/g_m + R_s}$$

with the result that

$$A_d \equiv \frac{v_{od}}{v_{id}} = \frac{R_D \parallel (R_L/2)}{1/g_m + R_s} \tag{9.37}$$

Video Example VE 9.1 Design of a MOS Differential Amplifier

Design a MOS differential amplifier to operate from ±1-V supplies and dissipate no more than 1 mW in its equilibrium state. Select the value of V_{OV} so that the value of v_{id} that steers the current from one side of the pair to the other is 0.25 V. The differential voltage gain A_d is to be 10 V/V. Assume $k_n' = 400\ \mu\text{A/V}^2$ and neglect the Early effect. Specify the required values of I, R_D, and W/L.

VE 9.1

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 9.15

EXERCISE

9.4 A MOS differential amplifier is operated at a total current of 0.8 mA, using transistors with a W/L ratio of 100, $\mu_n C_{ox} = 0.2\ \text{mA/V}^2$, $V_A = 20$ V, and $R_D = 5$ kΩ. Find V_{OV}, g_m, r_o, and A_d.
Ans. 0.2 V; 4 mA/V; 50 kΩ; 18.2 V/V

9.1.5 The Differential Amplifier with Current-Source Loads

To obtain higher gain, the passive resistances R_D can be replaced with current sources, as shown in Fig. 9.12(a). Here the current sources are realized with PMOS transistors Q_3 and Q_4, and V_G is a dc bias voltage that ensures that Q_3 and Q_4 each conducts a current equal to $I/2$. The differential voltage gain A_d can be found from the differential half-circuit shown in

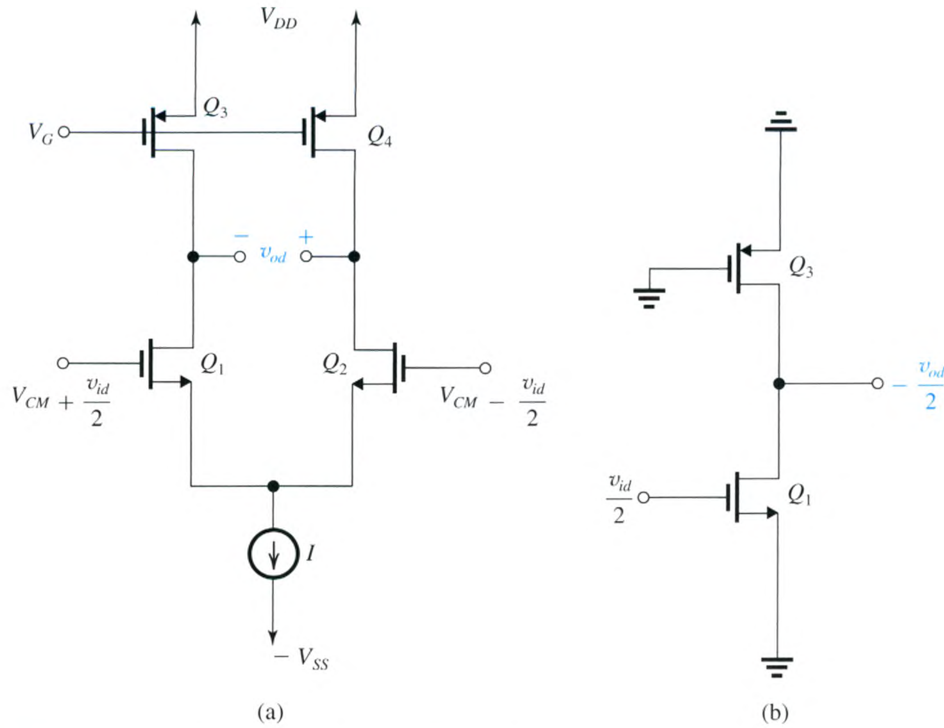


Figure 9.12 (**a**) Differential amplifier with current-source loads formed by Q_3 and Q_4. (**b**) Differential half-circuit of the amplifier in (**a**).

Fig. 9.12(b) as

$$A_d \equiv \frac{v_{od}}{v_{id}} = g_{m1}(r_{o1} \| r_{o3})$$

EXERCISE

9.5 The differential amplifier of Fig. 9.12(a) is fabricated in a 0.18-μm CMOS technology for which $\mu_n C_{ox} = 4\mu_p C_{ox} = 400\ \mu\text{A/V}^2$, $|V_t| = 0.5$ V, and $|V_A'| = 10$ V/μm. If the bias current $I = 200\ \mu$A and all transistors have a channel length twice the minimum and are operating at $|V_{OV}| = 0.2$ V, find W/L for each of Q_1, Q_2, Q_3, and Q_4, and determine the differential voltage gain A_d.
Ans. $(W/L)_{1,2} = 12.5$; $(W/L)_{3,4} = 50$; $A_d = 18$ V/V

THE LONG-TAILED PAIR

This idea using vacuum tubes was first documented by B. C. P. Matthews in 1934 in the *Proceedings of the Physical Society*, and was further developed by others in the late 1930s. The topology is simply that of a differential pair, where the term "long-tailed" refers to the biasing current source, which originally used a large-valued (hence long) resistor. Interestingly enough, the first application in measuring biological potentials in an electrically noisy environment continues to be an important one in modern medical instruments that utilize MOS devices.

9.1.6 Cascode Differential Amplifier

We can increase the gain of the differential amplifier using the cascode configuration studied in Section 8.5. Figure 9.13(a) shows a CMOS differential amplifier with cascoding applied to the amplifying transistors Q_1 and Q_2 via transistors Q_3 and Q_4, and to the current-source transistors Q_7 and Q_8 via transistors Q_5 and Q_6. The differential voltage gain can be found

Figure 9.13 (**a**) Cascode differential amplifier; and (**b**) its differential half-circuit.

from the differential half-circuit shown in Fig. 9.13(b) as

$$A_d \equiv \frac{v_{od}}{v_{id}} = g_{m1}(R_{on} \| R_{op}) \tag{9.38}$$

where

$$R_{on} = (g_{m3}r_{o3})r_{o1} \tag{9.39}$$

and,

$$R_{op} = (g_{m5}r_{o5})r_{o7} \tag{9.40}$$

EXERCISE

9.6 The CMOS cascode differential amplifier of Fig. 9.13(a) is fabricated in a 0.18-μm technology for which $\mu_n C_{ox} = 4\mu_p C_{ox} = 400\ \mu\text{A/V}^2$, $|V_t| = 0.5$ V, and $|V_A'| = 10$ V/μm. If the bias current $I = 200\ \mu$A, and all transistors have a channel length twice the minimum and are operating at $|V_{OV}| = 0.2$ V, find *W/L* for each of Q_1 to Q_8, and determine the differential voltage gain A_d.
Ans. $(W/L)_{1,2,3,4} = 12.5$; $(W/L)_{5,6,7,8} = 50$; $A_d = 648$ V/V

9.2 The BJT Differential Pair

Figure 9.14 shows the basic BJT differential-pair configuration. It is very similar to the MOSFET circuit and consists of two matched transistors, Q_1 and Q_2, whose emitters are joined together and biased by a constant-current source I. The latter is usually implemented

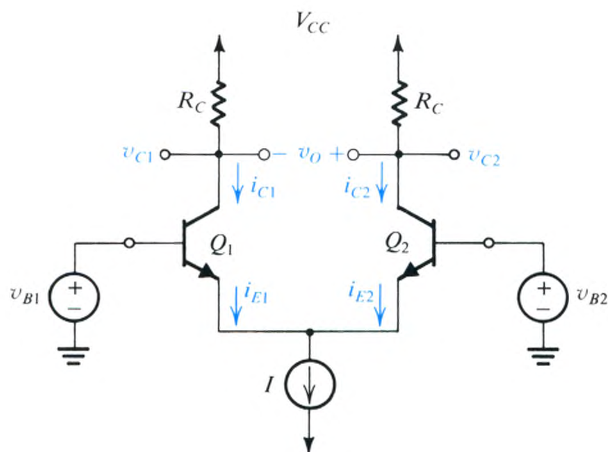


Figure 9.14 The basic BJT differential-pair configuration.

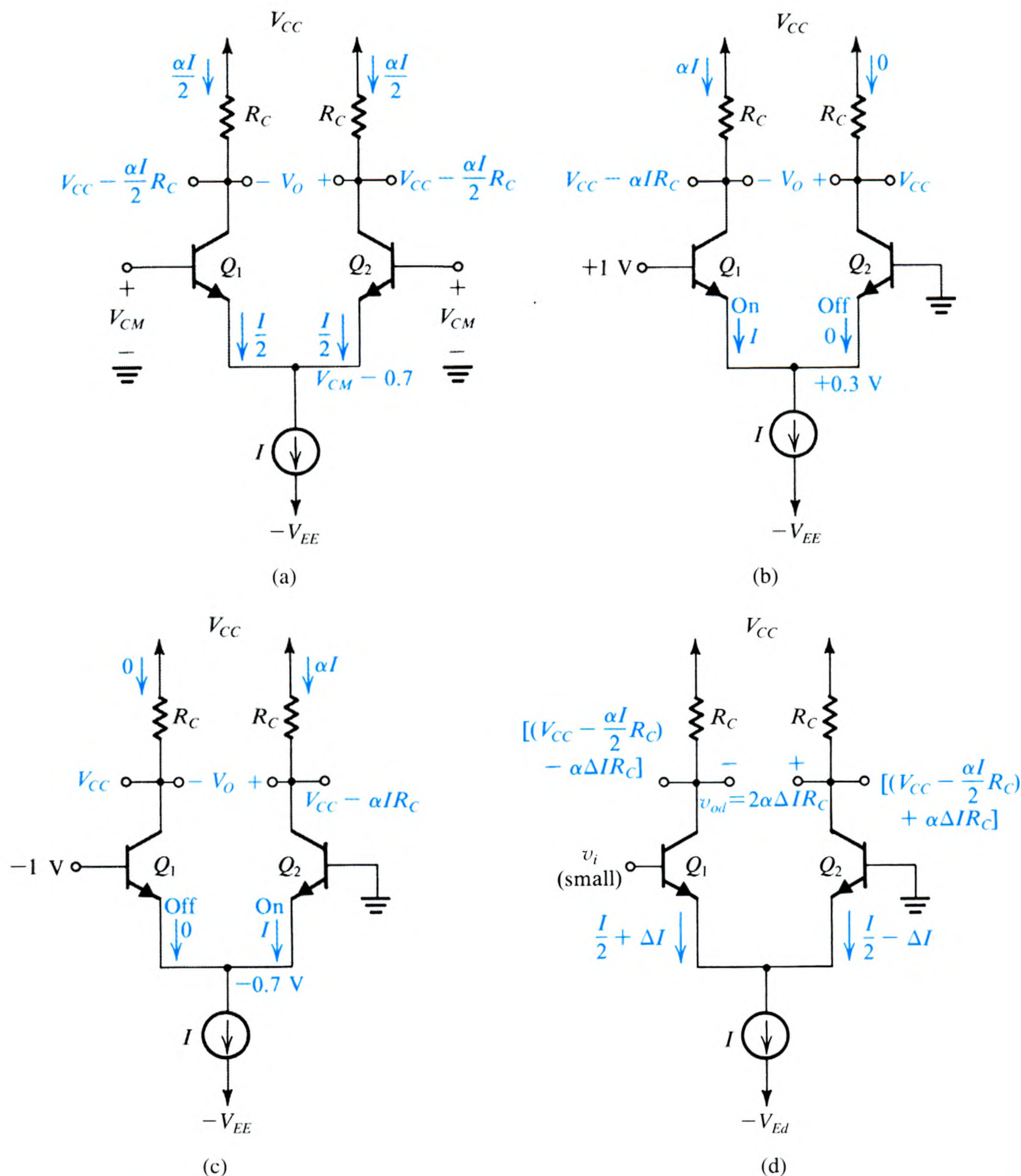


Figure 9.15 Different modes of operation of the BJT differential pair: **(a)** the differential pair with a common-mode input voltage V_{CM}; **(b)** the differential pair with a "large" differential input signal; **(c)** the differential pair with a large differential input signal of polarity opposite to that in **(b)**; **(d)** the differential pair with a small differential input signal v_i. Note that we have assumed the bias current source I to be ideal (i.e., it has an infinite output resistance) and thus I remains constant with the change in the voltage across it.

by a transistor circuit of the type studied in Sections 8.2 and 8.6. Although each collector is shown connected to the positive supply voltage V_{CC} through a resistance R_C, this connection is not essential to the operation of the differential pair—that is, in some applications the two collectors may be connected to current sources rather than resistive loads. It is essential, though, that the collector circuits be such that Q_1 and Q_2 operate in the active mode at all times.

9.2.1 Basic Operation

To see how the BJT differential pair works, consider first the case of a common-mode dc voltage V_{CM} applied to the two input terminals. That is, as shown in Fig. 9.15(a), $v_{B1} = v_{B2} = V_{CM}$. Since Q_1 and Q_2 are matched, and assuming an ideal bias current source I with infinite output resistance, it follows that the current I will remain constant and, from symmetry, that I will divide equally between the two devices. Thus $i_{E1} = i_{E2} = I/2$, and the voltage at the emitters will be $V_{CM} - V_{BE}$, where V_{BE} is the base–emitter voltage [assumed in Fig 9.15(a) to be approximately 0.7 V] corresponding to an emitter current of $I/2$. The voltage at each collector will be $V_{CC} - \frac{1}{2}\alpha IR_C$, and the difference in voltage between the two collectors, V_O, will be zero.

Now let us vary the value of the common-mode input voltage V_{CM}. As long as Q_1 and Q_2 remain in the active region, and the current source I has sufficient voltage across it to operate properly, the current I will still divide equally between Q_1 and Q_2, and the voltages at the collectors will not change. Thus, V_O remains equal to zero, indicating that the differential pair does not respond to (i.e., it *rejects*) changes in the common-mode input voltage.

As another experiment, let the voltage v_{B2} be set to a constant value, say, zero (by grounding B_2), and let $v_{B1} = +1$ V [see Fig. 9.15(b)]. With a bit of reasoning we can see that Q_1 will be on and conducting all of the current I and that Q_2 will be off. For Q_1 to be on (with $V_{BE1} = 0.7$ V), the emitter has to be at approximately +0.3 V, which keeps the EBJ of Q_2 reverse-biased. The collector voltages will be $v_{C1} = V_{CC} - \alpha IR_C$ and $v_{C2} = V_{CC}$, and $V_O = \alpha IR_C$.

Let us now change v_{B1} to –1 V [Fig. 9.15(c)]. Again with some reasoning we can see that Q_1 will turn off, and Q_2 will carry all the current I. The common emitter will be at –0.7 V, which means that the EBJ of Q_1 will be reverse biased by 0.3 V. The collector voltages will be $v_{C1} = V_{CC}$ and $v_{C2} = V_{CC} - \alpha IR_C$, and $V_O = -\alpha IR_C$.

From the foregoing, we see that the differential pair certainly responds to difference-mode (or differential) signals. In fact, with relatively small difference voltages we are able to steer the entire bias current from one side of the pair to the other. This current-steering property of the differential pair allows it to be used in logic circuits.

To use the BJT differential pair as a linear amplifier, we apply a very small differential signal (a few millivolts), which will result in one of the transistors conducting a current of $I/2 + \Delta I$; the current in the other transistor will be $I/2 - \Delta I$, with ΔI being proportional to the difference input voltage (see Fig. 9.15(d)). The output voltage taken between the two collectors, v_{od}, will be $2\alpha \Delta IR_C$, which is proportional to the differential input signal v_i. The small-signal operation of the differential pair will be studied shortly.

EXERCISE

9.7 Find v_E, v_{C1}, and v_{C2} in the circuit of Fig. E9.7. Assume that $|v_{BE}|$ of a conducting transistor is approximately 0.7 V and that $\alpha \simeq 1$.

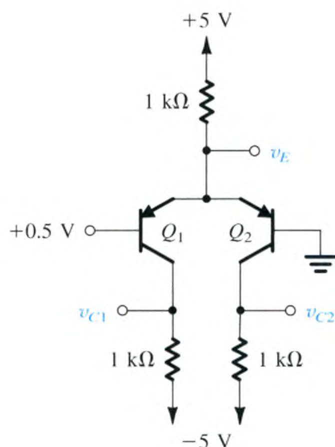


Figure E9.7

Ans. +0.7 V; −5 V; −0.7 V

9.2.2 Input Common-Mode Range

Refer to the circuit in Fig. 9.15(a). The allowable range of V_{CM} is determined at the upper end by Q_1 and Q_2 leaving the active mode and entering saturation. Thus

$$V_{CM\max} \simeq V_C + 0.4 = V_{CC} - \alpha \frac{I}{2} R_C + 0.4 \qquad (9.41)$$

The lower end of the V_{CM} range is determined by the need to provide a certain minimum voltage V_{CS} across the current source I to ensure its proper operation. Thus,

$$V_{CM\min} = -V_{EE} + V_{CS} + V_{BE} \qquad (9.42)$$

EXERCISE

9.8 Determine the input common-mode range for a bipolar differential amplifier operating from ±2.5-V power supplies and biased with a simple current source that delivers a constant current of 0.4 mA and requires a minimum of 0.3 V for its proper operation. The collector resistances $R_C = 5$ kΩ.
Ans. −1.5 V to +1.9 V

9.2.3 Large-Signal Operation

We now present a general analysis of the BJT differential pair of Fig. 9.14. If we denote the voltage at the common emitter by v_E and neglect the Early effect, the exponential relationship applied to each of the two transistors may be written

$$i_{E1} = \frac{I_S}{\alpha} e^{(v_{B1} - v_E)/V_T} \tag{9.43}$$

$$i_{E2} = \frac{I_S}{\alpha} e^{(v_{B2} - v_E)/V_T} \tag{9.44}$$

These two equations can be combined to obtain

$$\frac{i_{E1}}{i_{E2}} = e^{(v_{B1} - v_{B2})/V_T}$$

which can be manipulated to yield

$$\frac{i_{E1}}{i_{E1} + i_{E2}} = \frac{1}{1 + e^{(v_{B2} - v_{B1})/V_T}} \tag{9.45}$$

$$\frac{i_{E2}}{i_{E1} + i_{E2}} = \frac{1}{1 + e^{(v_{B1} - v_{B2})/V_T}} \tag{9.46}$$

The circuit imposes the additional constraint

$$i_{E1} + i_{E2} = I \tag{9.47}$$

Using Eq. (9.47) together with Eqs. (9.45) and (9.46) and substituting $v_{B1} - v_{B2} = v_{id}$ gives

$$i_{E1} = \frac{I}{1 + e^{-v_{id}/V_T}} \tag{9.48}$$

$$i_{E2} = \frac{I}{1 + e^{v_{id}/V_T}} \tag{9.49}$$

The collector currents i_{C1} and i_{C2} can be obtained simply by multiplying the emitter currents in Eqs. (9.48) and (9.49) by α, which is normally very close to unity.

The fundamental operation of the differential amplifier is illustrated by Eqs. (9.48) and (9.49). First, note that the amplifier responds only to the difference voltage v_{id}. That is, if $v_{B1} = v_{B2} = V_{CM}$, the current I divides equally between the two transistors irrespective of the value of the common-mode voltage V_{CM}. This is the essence of differential-amplifier operation, which also gives rise to its name.

Another important observation is that a relatively small difference voltage v_{id} will cause the current I to flow almost entirely in one of the two transistors. Figure 9.16 shows a plot of the two collector currents (assuming $\alpha \simeq 1$) as a function of the differential input signal. This normalized plot can be used universally. Observe that a difference voltage of about $4V_T$ ($\simeq$100 mV) is sufficient to switch the current almost entirely to one side of the BJT pair. Note that this is much smaller than the corresponding voltage for the MOS pair, $\sqrt{2}\, V_{OV}$. The fact that such a small signal can switch the current from one side of the BJT differential pair to the other means that the BJT differential pair can be used as a fast current switch.

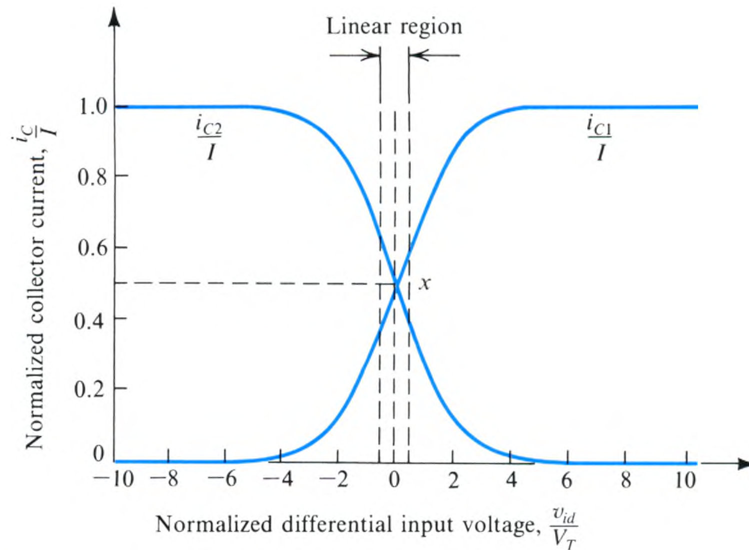


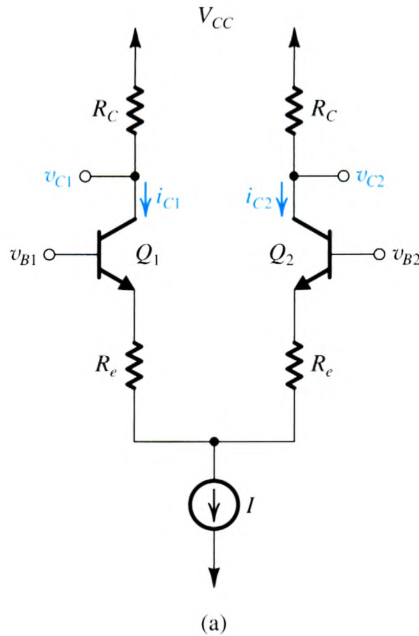
Figure 9.16 Transfer characteristics of the BJT differential pair of Fig. 9.14 assuming $\alpha \simeq 1$.

The nonlinear transfer characteristics of the differential pair, shown in Fig. 9.16, will not be utilized any further in this chapter. Rather, in the following we shall be interested specifically in the application of the differential pair as a small-signal amplifier. For this purpose, the difference input signal is limited to less than about $V_T/2$ in order that we may operate on a linear segment of the characteristics around the midpoint x (in Fig. 9.16).

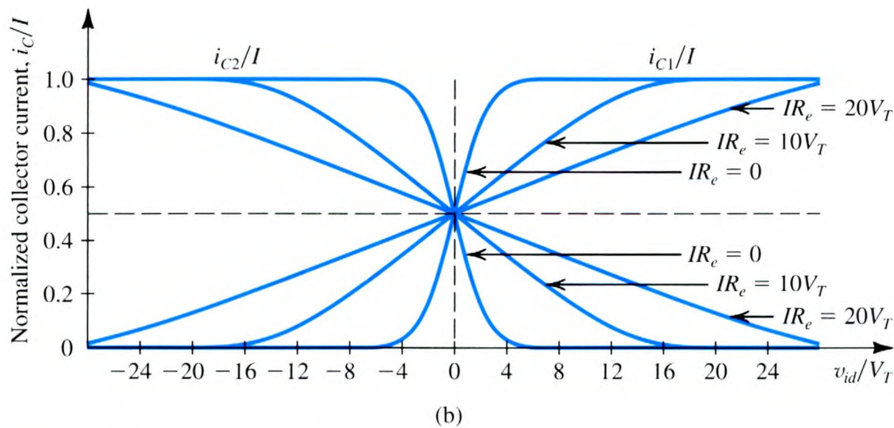
Before leaving the large-signal operation of the differential BJT pair, we wish to point out an effective technique frequently employed to extend the linear range of operation. It consists of including two equal resistances R_e in series with the emitters of Q_1 and Q_2, as shown in Fig. 9.17(a). The resulting transfer characteristics for three different values of R_e are sketched in Fig. 9.17(b). Observe that expansion of the linear range is obtained at the expense of reduced G_m (which is the slope of the transfer curve at $v_{id} = 0$) and hence reduced gain. This result should come as no surprise; R_e here is performing in exactly the same way as the emitter resistance R_e does in the CE amplifier with emitter degeneration (see Section 7.3.4). Finally, we also note that this linearization technique is in effect the bipolar counterpart of the technique employed for the MOS differential pair (Fig. 9.7). In the latter case, however, V_{OV} was varied by changing the transistors' W/L ratio, a design tool with no counterpart in the BJT.

EXERCISE

9.9 For the BJT differential pair of Fig. 9.14, find the value of input differential signal that is sufficient to cause $i_{E1} = 0.99I$.
Ans. 115 mV



(a)



(b)

Figure 9.17 The transfer characteristics of the BJT differential pair **(a)** can be linearized **(b)** (i.e., the linear range of operation can be extended) by including resistances in the emitters.

9.2.4 Small-Signal Operation

In this section we study the application of the BJT differential pair in small-signal amplification. Figure 9.18 shows the BJT differential pair with a difference voltage signal v_{id} applied between the two bases. Implied is that the dc level at the input—that is, the common-mode input voltage—has been somehow established. For instance, one of the two input terminals can be grounded and v_{id} applied to the other input terminal. Alternatively, the differential amplifier may be fed from the output of another differential amplifier. In the latter case, the voltage at one of the input terminals will be $V_{CM} + v_{id}/2$ while the voltage at the other input terminal will be $V_{CM} - v_{id}/2$.

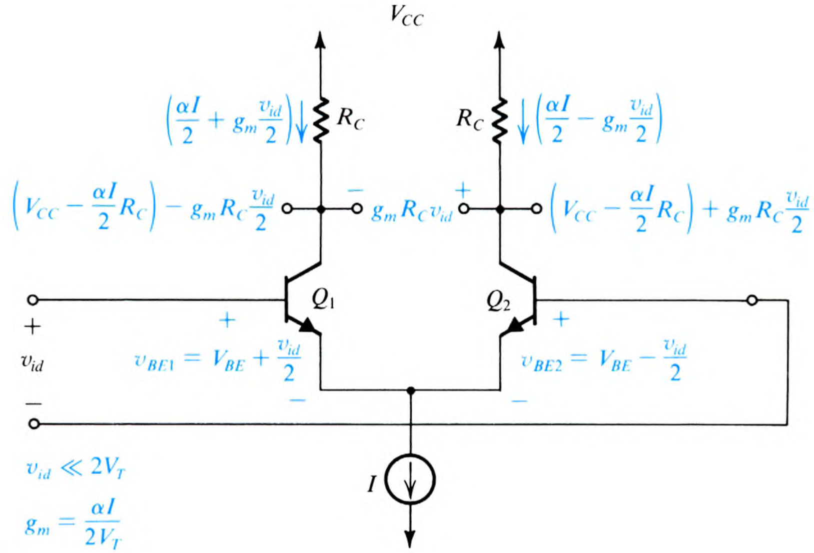


Figure 9.18 The currents and voltages in the differential amplifier when a small differential input signal v_{id} is applied.

The Collector Currents When v_{id} Is Applied For the circuit of Fig. 9.18, we may use Eqs. (9.48) and (9.49) to write

$$i_{C1} = \frac{\alpha I}{1 + e^{-v_{id}/V_T}} \tag{9.50}$$

$$i_{C2} = \frac{\alpha I}{1 + e^{v_{id}/V_T}} \tag{9.51}$$

Multiplying the numerator and the denominator of the right-hand side of Eq. (9.50) by $e^{v_{id}/2V_T}$ gives

$$i_{C1} = \frac{\alpha I e^{v_{id}/2V_T}}{e^{v_{id}/2V_T} + e^{-v_{id}/2V_T}}$$

Assume that $v_{id} \ll 2V_T$. We may thus expand the exponential $e^{\pm v_{id}/2V_T}$ in a series and retain only the first two terms:

$$i_{C1} \simeq \frac{\alpha I(1 + v_{id}/2V_T)}{1 + v_{id}/2V_T + 1 - v_{id}/2V_T}$$

Thus

$$i_{C1} = \frac{\alpha I}{2} + \frac{\alpha I}{2V_T}\frac{v_{id}}{2} \tag{9.52}$$

Similar manipulations can be applied to Eq. (9.51) to obtain

$$i_{C2} = \frac{\alpha I}{2} - \frac{\alpha I}{2V_T}\frac{v_{id}}{2} \tag{9.53}$$

Equations (9.52) and (9.53) tell us that when $v_{id} = 0$, the bias current I divides equally between the two transistors of the pair. Thus each transistor is biased at an emitter current of $I/2$. When

a "small-signal" v_{id} is applied differentially (i.e., between the two bases), the collector current of Q_1 increases by an increment i_c and that of Q_2 decreases by an equal amount. This ensures that the sum of the total currents in Q_1 and Q_2 remains constant, as constrained by the current-source bias. The incremental (or signal) current component i_c is given by

$$i_c = \frac{\alpha I}{2V_T}\frac{v_{id}}{2} \tag{9.54}$$

Equation (9.54) has an easy interpretation. First, note from the symmetry of the circuit (Fig. 9.18) that the differential signal v_{id} should divide equally between the base–emitter junctions of the two transistors. Thus the total base–emitter voltages will be

$$v_{BE}\big|_{Q1} = V_{BE} + \frac{v_{id}}{2}$$

$$v_{BE}\big|_{Q2} = V_{BE} - \frac{v_{id}}{2}$$

where V_{BE} is the dc BE voltage corresponding to an emitter current of $I/2$. Therefore, the collector current of Q_1 will increase by $g_m v_{id}/2$ and the collector current of Q_2 will decrease by $g_m v_{id}/2$. Here g_m denotes the transconductance of Q_1 and of Q_2, which are equal and given by

$$g_m = \frac{I_C}{V_T} = \frac{\alpha I/2}{V_T} \tag{9.55}$$

Thus Eq. (9.54) simply states that $i_c = g_m v_{id}/2$.

An Alternative Viewpoint There is an extremely useful alternative interpretation of the results above. Assume the current source I to be ideal. Its incremental resistance then will be infinite. Thus the voltage v_{id} appears across a total resistance of $2r_e$, where

$$r_e = \frac{V_T}{I_E} = \frac{V_T}{I/2} \tag{9.56}$$

Correspondingly, there will be a signal current i_e, as illustrated in Fig. 9.19, given by

$$i_e = \frac{v_{id}}{2r_e} \tag{9.57}$$

Thus the collector of Q_1 will exhibit a current increment i_c and the collector of Q_2 will exhibit a current decrement i_c:

$$i_c = \alpha i_e = \frac{\alpha v_{id}}{2r_e} = g_m \frac{v_{id}}{2} \tag{9.58}$$

Note that in Fig. 9.19(a) we have shown signal quantities only. It is implied, of course, that each transistor is biased at an emitter current of $I/2$. For greater emphasis, we show in Fig. 9.19(b) the equivalent circuit obtained by replacing each BJT with its T model.

This method of analysis is particularly useful when resistances are included in the emitters, as shown in Fig. 9.20. For this circuit we have

$$i_e = \frac{v_{id}}{2r_e + 2R_e} \tag{9.59}$$

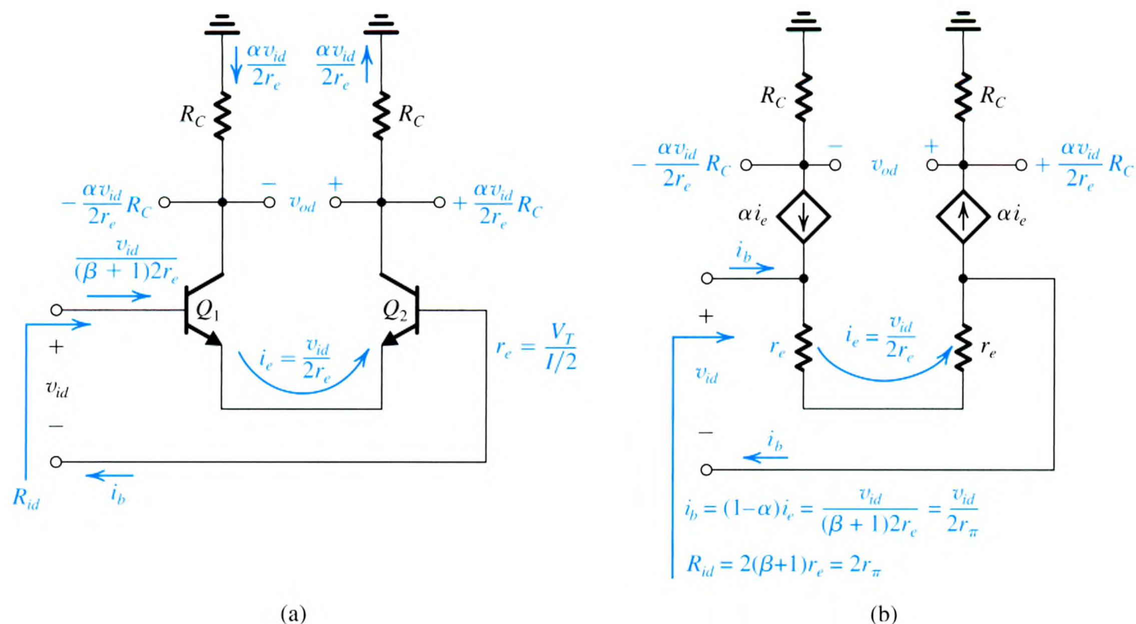


Figure 9.19 A simple technique for determining the signal currents in a differential amplifier excited by a differential voltage signal v_{id}; dc quantities are not shown. While Fig. 9.19(a) utilizes the BJT T model implicitly, the T model of both BJTs are shown explicitly in Fig. 9.19(b).

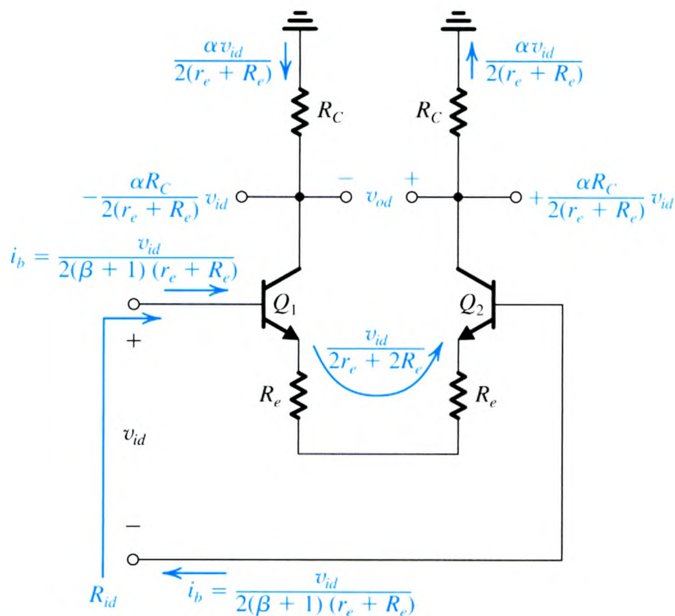


Figure 9.20 A differential amplifier with emitter resistances. Only signal quantities are shown (in color).

Input Differential Resistance Unlike the MOS differential amplifier, which has an infinite input resistance, the bipolar differential pair exhibits a finite input resistance, a result of the finite β of the BJT.

The input differential resistance is the resistance seen between the two bases; that is, it is the resistance seen by the differential input signal v_{id}. For the differential amplifier in Figs. 9.18 and 9.19 we can see that the base current of Q_1 shows an increment i_b and the base current of Q_2 shows an equal decrement,

$$i_b = \frac{i_e}{\beta+1} = \frac{v_{id}/2r_e}{\beta+1} \tag{9.60}$$

Thus the differential input resistance R_{id} is given by

$$R_{id} \equiv \frac{v_{id}}{i_b} = (\beta+1)2r_e = 2r_\pi \tag{9.61}$$

This result is just a restatement of the familiar resistance-reflection rule: namely, *the resistance between the two bases is equal to the total resistance in the emitter circuit multiplied by* $(\beta + 1)$. We can use this rule to find the input differential resistance for the circuit in Fig. 9.20:

$$R_{id} = (\beta+1)(2r_e + 2R_e) \tag{9.62}$$

Differential Voltage Gain We have established that for small difference input voltages ($v_{id} \ll 2\,V_T$; i.e., v_{id} smaller than about 20 mV), the collector currents are given by

$$i_{C1} = I_C + g_m \frac{v_{id}}{2} \tag{9.63}$$

$$i_{C2} = I_C - g_m \frac{v_{id}}{2} \tag{9.64}$$

where

$$I_C = \frac{\alpha I}{2} \tag{9.65}$$

Thus the total voltages at the collectors will be

$$v_{C1} = (V_{CC} - I_C R_C) - g_m R_C \frac{v_{id}}{2} \tag{9.66}$$

$$v_{C2} = (V_{CC} - I_C R_C) + g_m R_C \frac{v_{id}}{2} \tag{9.67}$$

The quantities in parentheses are the dc voltages at each of the two collectors.

As in the MOS case, the output voltage signal of a bipolar differential amplifier can be taken *differentially* (i.e., between the two collectors, $v_{od} = v_{c2} - v_{c1}$). The differential gain of the differential amplifier will be

$$A_d = \frac{v_{od}}{v_{id}} = g_m R_C \tag{9.68}$$

For the differential amplifier with resistances in the emitter leads (Fig. 9.20), the differential gain is given by

$$A_d = \frac{\alpha(2R_C)}{2r_e + 2R_e} \simeq \frac{R_C}{r_e + R_e} \tag{9.69}$$

This equation is a familiar one: It states that *the voltage gain is equal to the ratio of the total resistance in the collector circuit* ($2R_C$) *to the total resistance in the emitter circuit* ($2r_e + 2R_e$).

The Differential Half-Circuit As in the MOS case, we can find the differential gain of the BJT differential amplifier by considering its differential half-circuit. Figure 9.21(a) shows a differential amplifier fed by a differential signal v_{id} that is applied in a **complementary** (**push–pull** or **balanced**) manner. That is, while the base of Q_1 is raised by $v_{id}/2$, the base of Q_2 is lowered by $v_{id}/2$. We have also included the output resistance R_{EE} of the bias current source. From symmetry, it follows that the signal voltage at the emitters will be zero. Thus the circuit is equivalent to the two common-emitter amplifiers shown in Fig. 9.21(b), where each of the two transistors is biased at an emitter current of $I/2$. Note that the finite output resistance R_{EE} of the current source will have no effect on the operation. The equivalent circuit in Fig. 9.21(b) is valid for differential operation only.

Since in Fig. 9.21, $v_{o2} = -v_{o1} = v_{od}/2$, the two common-emitter transistors in Fig. 9.21(b) yield similar results about the performance of the differential amplifier. Thus only one is needed to analyze the differential small-signal operation of the differential amplifier, and it is known as the **differential half-circuit**. If we take the common-emitter transistor fed with $+v_{id}/2$ as the differential half-circuit and replace the transistor with its low-frequency, hybrid-π, equivalent-circuit model, we will get the circuit in Fig. 9.22. In evaluating the model parameters r_π, g_m, and r_o, we must recall that the half-circuit is biased at $I/2$. The voltage gain of the differential amplifier is equal to the voltage gain of the half-circuit—that is, $v_{o1}/(v_{id}/2)$. Here, we note that including r_o will modify the gain expression in Eq. (9.68) to

$$A_d = g_m(R_C \parallel r_o) \tag{9.70}$$

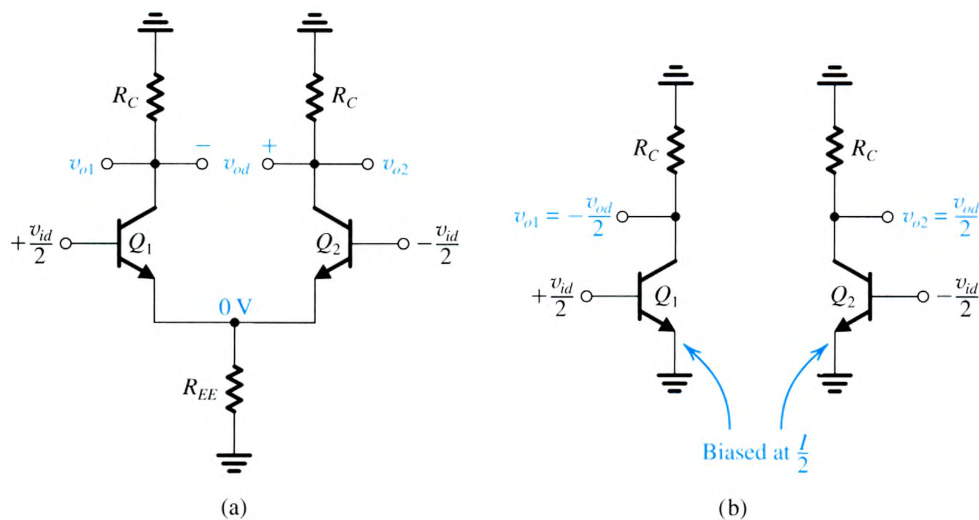


Figure 9.21 Equivalence of the BJT differential amplifier in **(a)** to the two common-emitter amplifiers in **(b)**. This equivalence applies only for differential input signals. Either of the two common-emitter amplifiers in **(b)** can be used to find the differential gain, differential input resistance, frequency response, and so on, of the differential amplifier.

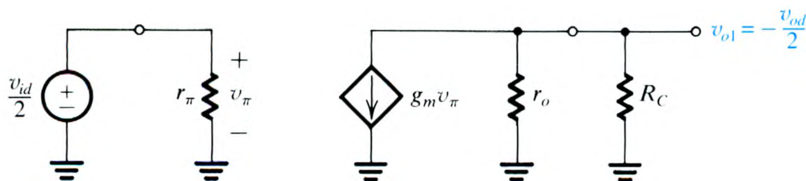


Figure 9.22 Equivalent-circuit model of the differential half-circuit formed by Q_1 in Fig. 9.21(b).

The input differential resistance of the differential amplifier is twice that of the half-circuit—that is, $2r_\pi$. Finally, we note that the differential half-circuit of the amplifier of Fig. 9.20 is a common-emitter transistor with a resistance R_e in the emitter lead.

Example 9.3

The differential amplifier in Fig. 9.23 uses transistors with $\beta = 100$. Evaluate the following:

(a) The input differential resistance R_{id}.
(b) The overall differential voltage gain v_{od}/v_{sig} (neglect the effect of r_o).

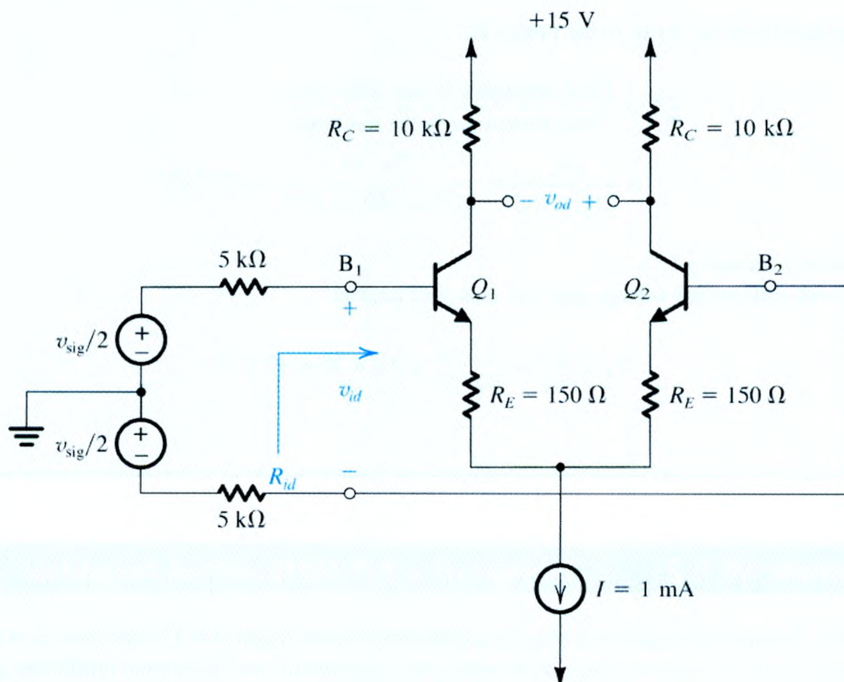


Figure 9.23 Circuit for Example 9.3.

Example 9.3 *continued*

Solution

(a) Each transistor is biased at an emitter current of 0.5 mA. Thus

$$r_{e1} = r_{e2} = \frac{V_T}{I_E} = \frac{25 \text{ mV}}{0.5 \text{ mA}} = 50\ \Omega$$

The input differential resistance can now be found as

$$\begin{aligned} R_{id} &= 2(\beta+1)(r_e + R_E) \\ &= 2 \times 101 \times (50 + 150) \simeq 40 \text{ k}\Omega \end{aligned}$$

(b) The voltage gain from the signal source to the bases of Q_1 and Q_2 is

$$\begin{aligned} \frac{v_{id}}{v_{sig}} &= \frac{R_{id}}{R_{sig} + R_{id}} \\ &= \frac{40}{5+5+40} = 0.8 \text{ V/V} \end{aligned}$$

The voltage gain from the bases to the output is

$$\begin{aligned} \frac{v_{od}}{v_{id}} &\simeq \frac{\text{Total resistance in the collectors}}{\text{Total resistance in the emitters}} \\ &= \frac{2R_C}{2(r_e + R_E)} = \frac{2 \times 10}{2(50+150) \times 10^{-3}} = 50 \text{ V/V} \end{aligned}$$

where we have assumed $\alpha \simeq 1$.

The overall differential voltage gain can now be found as

$$A_d = \frac{v_{od}}{v_{sig}} = \frac{v_{id}}{v_{sig}} \frac{v_{od}}{v_{id}} = 0.8 \times 50 = 40 \text{ V/V}$$

Video Example VE 9.2 Design of a BJT Differential Amplifier

Design a BJT differential amplifier to amplify a differential input signal of 0.1 V and provide a differential output signal of 2 V. To ensure adequate linearity, it is required to limit the signal amplitude across each base–emitter junction to a maximum of 5 mV. Another design requirement is that the differential input resistance must be at least 100 kΩ. The BJTs available are specified to have $\beta \geq 100$. Give the circuit configuration and specify the values of all its components.

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.


VE 9.2

Related end-of-chapter problem: **9.42**

EXERCISE

9.10 For the circuit in Fig. 9.18, let $I = 1$ mA, $V_{CC} = 15$ V, $R_C = 10$ kΩ, with $\alpha = 1$, and let the input voltages be $v_{B1} = 5 + 0.005 \sin 2\pi \times 1000t$, volts, and $v_{B2} = 5 - 0.005 \sin 2\pi \times 1000t$, volts. (a) If the BJTs are specified to have v_{BE} of 0.7 V at a collector current of 1 mA, find the voltage at the emitters. (b) Find g_m for each of the two transistors. (c) Find i_C for each of the two transistors. (d) Find v_C for each of the two transistors. (e) Find the voltage between the two collectors. (f) Find the gain experienced by the 1000-Hz signal.

Ans. (a) 4.317 V; (b) 20 mA/V; (c) $i_{C1} = 0.5 + 0.1 \sin 2\pi \times 1000t$, mA and $i_{C2} = 0.5 - 0.1 \sin 2\pi \times 1000t$, mA; (d) $v_{C1} = 10 - 1 \sin 2\pi \times 1000t$, V and $v_{C2} = 10 + 1 \sin 2\pi \times 1000t$, V; (e) $v_{C2} - v_{C1} = 2 \sin 2\pi \times 1000t$, V; (f) 200 V/V

9.3 Common-Mode Rejection

Thus far, we have seen that the differential amplifier responds to a differential input signal and completely rejects a common-mode signal. This latter point was shown clearly both at the outset of our discussion of differential amplifiers and in Example 9.1, where we saw that changes in V_{CM} over a wide range resulted in no change in the voltage at either of the two drains. The same phenomenon was demonstrated for the BJT differential amplifier in Section 9.2.1. This highly desirable result is, however, a consequence of our assumption that the current source that supplies the bias current I is ideal. We will now consider the situation when a more realistic current source with a finite output resistance is used.

9.3.1 The MOS Case

Figure 9.24(a) shows a MOS differential amplifier biased with a current source having an output resistance R_{SS}. As before, the dc voltage at the input is defined by V_{CM}. Here, however, we also have an incremental signal v_{icm} applied to both input terminals. This *common-mode input signal* can represent an interference signal or noise that is picked up by both inputs and is clearly undesirable. Our objective now is to find how much of v_{icm} makes its way to the output of the amplifier.

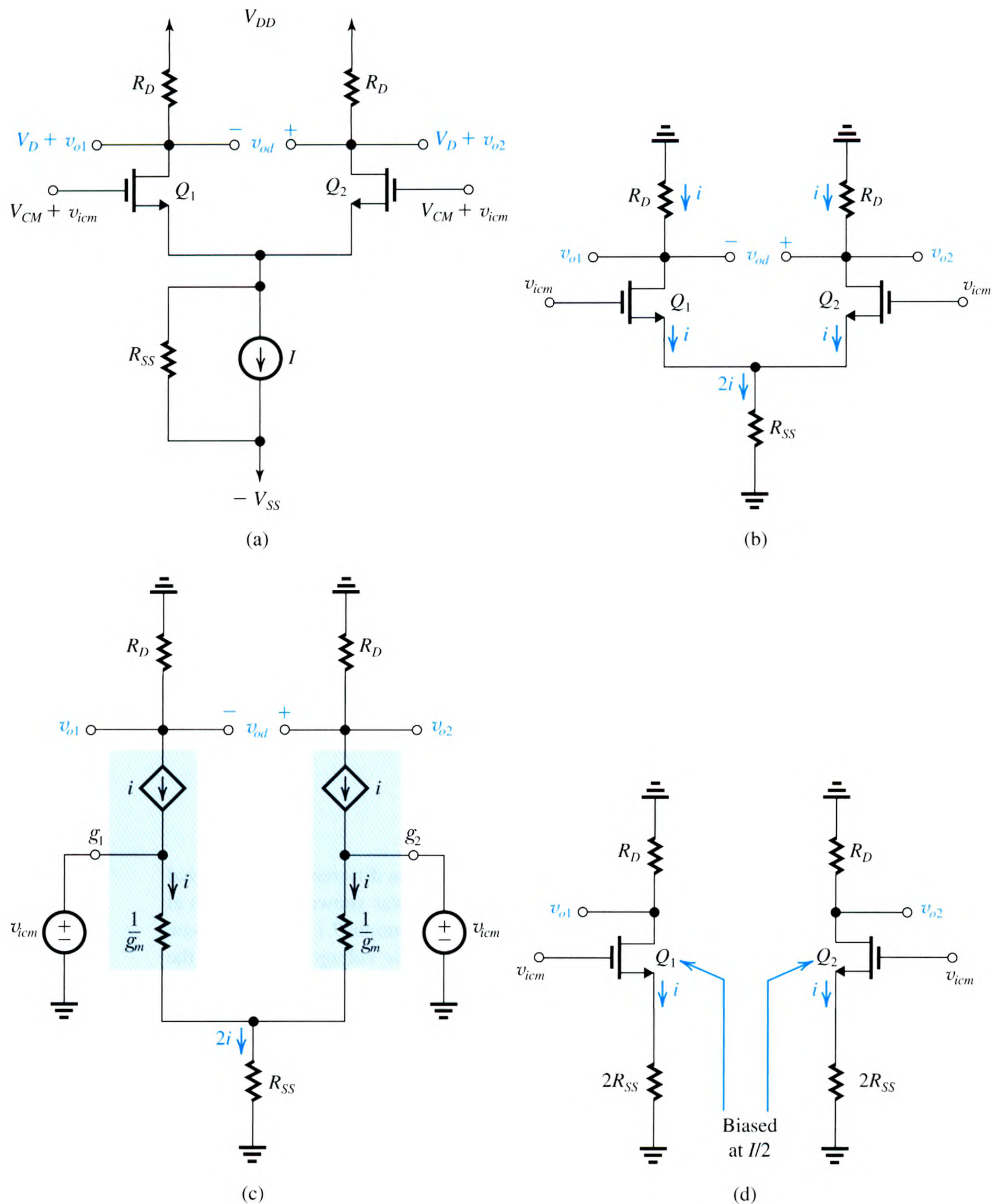


Figure 9.24 (**a**) A MOS differential amplifier with a common-mode input signal v_{icm} superimposed on the input dc common-mode voltage V_{CM}. (**b**) The amplifier circuit prepared for small-signal analysis. (**c**) The amplifier circuit with the transistors replaced with their T model and r_o neglected. (**d**) The circuit in (**b**) split into its two halves; each half is said to be a "CM half-circuit."

Before we determine the common-mode gain of the amplifier, we need to address the effect R_{SS} has on the bias current of Q_1 and Q_2. That is, with v_{icm} set to zero, the bias current in Q_1 and Q_2 will no longer be $I/2$ but will be larger than $I/2$ by an amount determined by V_{CM} and R_{SS}. However, since R_{SS} is usually very large, this additional dc current in each of Q_1 and Q_2 is usually small enough that we can neglect it, thus assuming that Q_1 and Q_2 continue to operate at a bias current of $I/2$. You might also be wondering about the effect of R_{SS} on the differential gain. The answer is very simple: The virtual ground that develops on the common-source terminal when a differential input signal is applied results in a zero signal current through R_{SS}; hence R_{SS} has no effect on the value of A_d.

To determine the response of the differential amplifier to the common-mode input signal v_{icm}, consider the circuit in Fig. 9.24(b), where we have replaced each of V_{DD} and V_{SS} by a short circuit and I by an open circuit. The circuit is obviously symmetrical, and thus the two transistors will carry equal signal currents, denoted i. The value of i can be easily determined by replacing each of Q_1 and Q_2 with its T model and, for simplicity, neglecting r_o. The resulting equivalent circuit is shown in Fig. 9.24(c), from which we can write

$$v_{icm} = \frac{i}{g_m} + 2iR_{SS} \tag{9.71}$$

Thus,

$$i = \frac{v_{icm}}{1/g_m + 2R_{SS}} \tag{9.72}$$

The voltages at the drain of Q_1 and Q_2 can now be found as

$$v_{o1} = v_{o2} = -R_D i$$

resulting in

$$v_{o1} = v_{o2} = -\frac{R_D}{1/g_m + 2R_{SS}} v_{icm} \tag{9.73}$$

It follows that both v_{o1} and v_{o2} will be corrupted by the common-mode signal v_{icm} and will be given approximately by

$$\frac{v_{o1}}{v_{icm}} = \frac{v_{o2}}{v_{icm}} \simeq -\frac{R_D}{2R_{SS}} \tag{9.74}$$

where we have assumed that $2R_{SS} \gg 1/g_m$. Nevertheless, because $v_{o1} = v_{o2}$, the differential output voltage v_{od} will remain free of common-mode interference:

$$v_{od} = v_{o2} - v_{o1} = 0 \tag{9.75}$$

Thus the circuit still rejects common-mode signals! Unfortunately, however, this will not be the case if the circuit is not perfectly matched, as we will now show.

Before we proceed, notice that we can obtain all the results above by considering only half the differential amplifier. Figure 9.24(d) shows the two half-circuits of the differential amplifier that apply for common-mode analysis. To see the equivalence, observe that each of the two half-circuits indeed carries a current i given by Eq. (9.72) and the voltages at the source terminals are equal ($v_s = 2iR_{SS}$). Thus the two sources can be joined, returning the circuit to the original form in Fig. 9.24(b). Each of the circuits in Fig. 9.24(d) is known as the **common-mode half-circuit**. Note the difference between the CM half-circuit and the differential half-circuit.

Effect of R_D Mismatch When the two drain resistances exhibit a mismatch ΔR_D, as they inevitably do, the voltages at the two drains will no longer have equal components due to v_{icm}. Rather, if the load of Q_1 is R_D and that of Q_2 is $(R_D + \Delta R_D)$ the drain signal voltages arising from v_{icm} will be

$$v_{o1} \simeq -\frac{R_D}{2R_{SS}} v_{icm} \tag{9.76}$$

and

$$v_{o2} \simeq -\frac{R_D + \Delta R_D}{2R_{SS}} v_{icm} \tag{9.77}$$

Thus,

$$v_{od} = v_{o2} - v_{o1} = -\frac{\Delta R_D}{2R_{SS}} v_{icm} \tag{9.78}$$

and we can find the **common-mode gain** A_{cm} as

$$A_{cm} \equiv \frac{v_{od}}{v_{icm}} = -\frac{\Delta R_D}{2R_{SS}} \tag{9.79}$$

which can be expressed in the alternate form

$$A_{cm} = -\left(\frac{R_D}{2R_{SS}}\right)\left(\frac{\Delta R_D}{R_D}\right) \tag{9.80}$$

It follows that a mismatch in the drain resistances causes the differential amplifier to have a finite common-mode gain. Thus, a portion of the interference or noise signal v_{icm} will appear as a component of v_{od}. A measure of the effectiveness of the differential amplifier in amplifying differential-mode signals and rejecting common-mode interference is the ratio of the magnitude of its differential gain $|A_d|$ to the magnitude of its common-mode gain $|A_{cm}|$. This ratio is termed **common-mode rejection ratio (CMRR)**. Thus,

$$\text{CMRR} \equiv \frac{|A_d|}{|A_{cm}|} \tag{9.81}$$

and is usually expressed in decibels,

$$\text{CMRR(dB)} = 20 \log \frac{|A_d|}{|A_{cm}|} \tag{9.82}$$

For the case of a MOS differential amplifier with drain resistances R_D that exhibit a mismatch ΔR_D, the CMRR can be found as the ratio of A_d in Eq. (9.35) to A_{cm} in Eq. (9.79), thus

$$\text{CMRR} = (2g_m R_{SS}) \Big/ \left(\frac{\Delta R_D}{R_D}\right) \tag{9.83}$$

It follows that to obtain a high CMRR, we should utilize a bias current source with a high output resistance R_{SS}, and we should strive to obtain a high degree of matching between the drain resistances (i.e., keep $\Delta R_D/R_D$ small).

EXERCISE

9.11 A MOS differential pair operated at a bias current of 0.8 mA employs transistors with $W/L = 100$ and $\mu_n C_{ox} = 0.2\ \text{mA/V}^2$, using $R_D = 5\ \text{k}\Omega$ and $R_{SS} = 25\ \text{k}\Omega$. Find the differential gain, the common-mode gain when the drain resistances have a 1% mismatch, and the CMRR.
Ans. 20 V/V; 0.001 V/V; 86 dB

Effect of g_m Mismatch on CMRR Another possible mismatch between the two halves of the MOS differential pair is a mismatch in g_m of the two transistors. If the two transistors Q_1 and Q_2 have a mismatch Δg_m in their g_m values, this mismatch leads to a finite common-mode gain given by

$$A_{cm} \simeq \left(\frac{R_D}{2R_{SS}}\right)\left(\frac{\Delta g_m}{g_m}\right) \tag{9.84}$$

The corresponding CMRR will be

$$\text{CMRR} = (2g_m R_{SS}) \bigg/ \left(\frac{\Delta g_m}{g_m}\right) \tag{9.85}$$

Note that both expressions have exactly the same form as the corresponding expressions for the case of R_D mismatch. Thus, as in that case, to keep CMRR high, we use a biasing current source with a high output resistance R_{SS} and, of course, strive to maintain a high degree of matching between Q_1 and Q_2.

EXERCISE

9.12 For the MOS amplifier specified in Exercise 9.11, compute the CMRR resulting from a 1% mismatch in g_m.
Ans. 86 dB

Example 9.4

In this example we consider the design of the current source that supplies the bias current of a MOS differential amplifier. Imagine we have to achieve a CMRR of 100 dB and assume that the only source of mismatch between Q_1 and Q_2 is a 2% mismatch in their W/L ratios. Let $I = 200\ \mu\text{A}$ and assume

Example 9.4 *continued*

that all transistors are to be operated at $V_{OV} = 0.2$ V. For the 0.18-μm CMOS fabrication process available, $V_A' = 5$ V/μm. If a simple current source is used for I, what channel length is required? If a cascode current source is used, what channel length is needed for the two transistors in the cascode?

Solution

A mismatch in W/L results in a g_m mismatch that can be found from the expression of g_m:

$$g_m = (\mu_n C_{ox})\left(\frac{W}{L}\right)V_{OV} \qquad (9.86)$$

We can see that an error of 2% in W/L will result in an error in g_m of 2%. That is, the 2% mismatch in the W/L ratios of Q_1 and Q_2 will result in a 2% mismatch in their g_m values. The resulting CMRR can be found from Eq. (9.85), repeated here:

$$\text{CMRR} = (2g_m R_{SS})\Big/\left(\frac{\Delta g_m}{g_m}\right)$$

Now, a 100-dB CMRR corresponds to a ratio of 10^5; thus,

$$10^5 = (2g_m R_{SS})/0.02 \qquad (9.87)$$

The value of g_m can be found from

$$g_m = \frac{2I_D}{V_{OV}} = \frac{2 \times (I/2)}{V_{OV}}$$

$$= \frac{2 \times 0.1}{0.2} = 1 \text{ mA/V}$$

Substituting in Eq. (9.87) gives

$$R_{SS} = 1 \text{ M}\Omega$$

Now if the current source is implemented with a single transistor, its r_o must be

$$r_o = R_{SS} = 1 \text{ M}\Omega$$

Thus,

$$\frac{V_A}{I} = 1 \text{ M}\Omega$$

Substituting $I = 200$ μA, we find the required value of V_A as

$$V_A = 200 \text{ V}$$

Since $V_A = V_A'L = 5L$, the required value of L will be

$$L = 40 \text{ μm}$$

which is very large!

Using a cascode current source, we have

$$R_{SS} = (g_m r_o) r_o$$

where

$$g_m = \frac{2I}{V_{OV}} = \frac{2 \times 0.2}{0.2} = 2 \text{ mA/V}$$

Thus,

$$1000 = 2 \times r_o^2$$

$$r_o = 22.36 \text{ k}\Omega$$

and the required V_A now becomes

$$22.36 = \frac{V_A}{I} = \frac{V_A}{0.2}$$

$$V_A = 4.47 \text{ V}$$

which implies a channel length for each of the two transistors in the cascode of

$$L = \frac{4.47}{V'_A} = \frac{4.47}{5} = 0.89 \ \mu\text{m}$$

a considerable reduction from the case of a simple current source, and indeed a practical value.

Video Example VE 9.3 Gain and CMRR of a MOS Differential Amplifier

An NMOS differential pair is biased by a current source $I = 0.2$ mA having an output resistance $R_{SS} = 100$ kΩ. The amplifier has drain resistances $R_D = 10$ kΩ, using transistors with $k_n = 3$ mA/V^2, and r_o that is large. If the output is taken differentially and there is a 1% mismatch between the drain resistances, find $|A_d|$, $|A_{cm}|$, and CMRR.

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

VE 9.3

Related end-of-chapter problem: 9.55

Differential versus Single-Ended Output The above study of common-mode rejection was predicated on the assumption that the output of the differential amplifier is taken differentially, that is, between the drains of Q_1 and Q_2. In some cases one might decide to take the output single-endedly; that is, between one of the drains and ground. If this is done,

the CMRR is reduced dramatically. We can see this quantitatively by observing that taking the output single-endedly results in the common-mode gain given by Eq. (9.74):

$$A_{cm} = -\frac{R_D}{2R_{SS}} \tag{9.88}$$

In this case, the differential gain is given by

$$A_d = -\frac{1}{2} g_m R_D \tag{9.89}$$

and the CMRR is

$$\text{CMRR} = g_m R_{SS} \tag{9.90}$$

which is much smaller than the CMRR for the case of differential output [Eqs. (9.83) and (9.85)].

We can conclude that to obtain a large CMRR, the output of the differential amplifier must be taken differentially. If a single-ended output is required, conversion from differential to single-ended must be done carefully to avoid loss of CMRR, as we will show in Section 9.5.

EXERCISE

9.13 For the MOS differential amplifier considered in Exercise 9.11, find the differential gain, the common-mode gain, and the CMRR obtained if the output is taken single-endedly.
Ans. 10 V/V; 0.1 V/V; 40 dB

9.3.2 The BJT Case

An identical development applies for studying the common-mode rejection of the BJT differential amplifier. Figure 9.25 shows a bipolar differential amplifier with an input common-mode signal v_{icm}. Here R_{EE} is the output resistance of the bias current source I. We wish to find the voltages that result from v_{icm} at the collectors of Q_1 and Q_2, v_{o1} and v_{o2}, and between the two collectors, v_{od}. To that end, we use the **common-mode half-circuits** shown in Fig. 9.25(b). The signal v_{o1} that appears at the collector of Q_1 in response to v_{icm} will be

$$v_{o1} = -\frac{\alpha R_C}{r_e + 2R_{EE}} v_{icm} \tag{9.91}$$

Similarly, v_{o2} will be

$$v_{o2} = -\frac{\alpha R_C}{r_e + 2R_{EE}} v_{icm} \tag{9.92}$$

where we have neglected the transistor r_o, for simplicity. We can find the differential output signal v_{od} as

$$v_{od} = v_{o2} - v_{o1} = 0$$

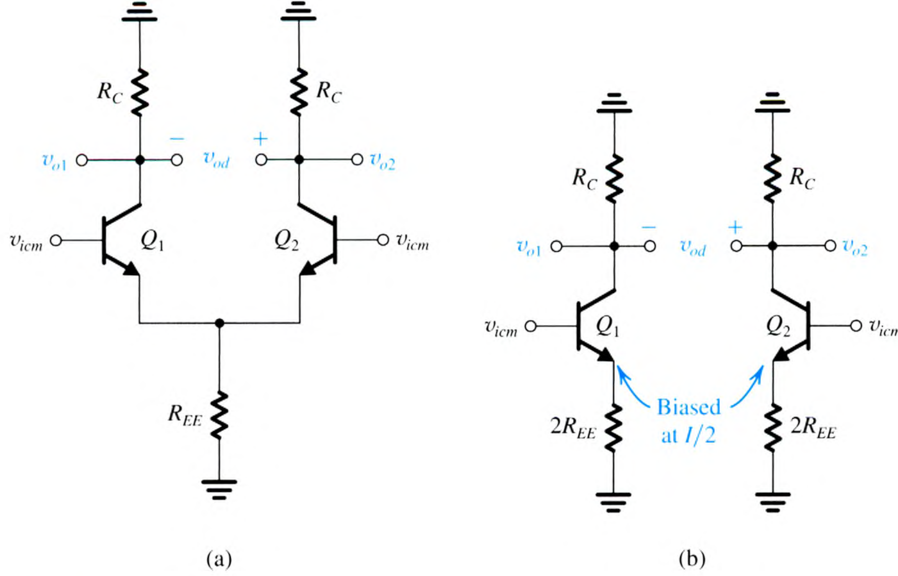


Figure 9.25 **(a)** The differential amplifier fed by a common-mode input signal v_{icm}. **(b)** Equivalent "half-circuits" for common-mode calculations.

Thus, while the voltages at the two collectors will contain common-mode noise or interference components, the output differential voltage will be free from such interference. This condition, however, is based on the assumption of perfect matching between the two sides of the differential amplifier. Any mismatch will result in v_{od} acquiring a component proportional to v_{icm}. For example, a mismatch ΔR_C between the two collector resistances results in

$$A_{cm} \equiv \frac{v_{od}}{v_{icm}} = -\frac{\alpha \Delta R_C}{2R_{EE} + r_e} \tag{9.93}$$

Since $\alpha \simeq 1$, $r_e \ll 2R_{EE}$, Eq. (9.93) can be approximated and written in the form

$$A_{cm} \simeq -\left(\frac{R_C}{2R_{EE}}\right)\left(\frac{\Delta R_C}{R_C}\right) \tag{9.94}$$

The common-mode rejection ratio can now be found from

$$\text{CMRR} = \frac{|A_d|}{|A_{cm}|}$$

together with using Eqs. (9.68) and (9.94), with the result that

$$\text{CMRR} = (2g_m R_{EE}) \bigg/ \left(\frac{\Delta R_C}{R_C}\right) \tag{9.95}$$

which is similar in form to the expression for the MOS pair [Eq. (9.83)]. Thus, to obtain a high CMRR, we design the current source to have a large output resistance R_{EE} and strive for close matching of the collector resistances.

Common-Mode Input Resistance The definition of the common-mode input resistance R_{icm} is illustrated in Fig. 9.26(a). Figure 9.26(b) shows the equivalent common-mode

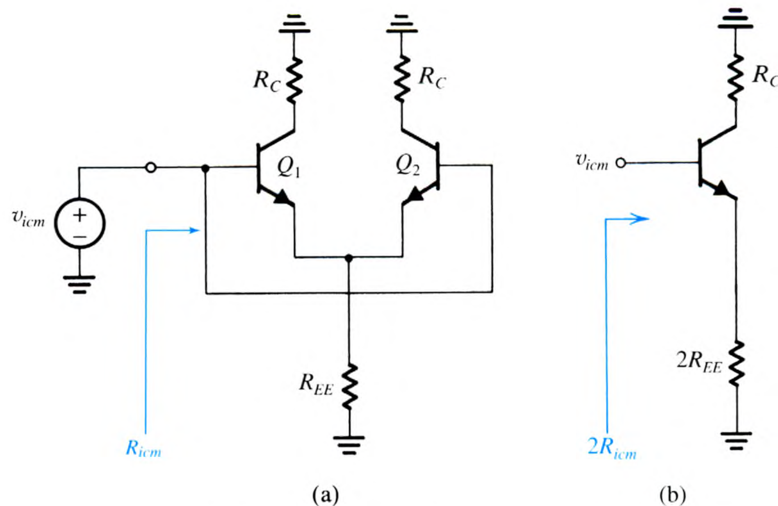


Figure 9.26 **(a)** Definition of the input common-mode resistance R_{icm}. **(b)** The equivalent common-mode half-circuit.

half-circuit; its input resistance is $2R_{icm}$. The value of $2R_{icm}$ can be determined by analyzing the circuit of Fig. 9.26(b) while taking r_o into account (because R_{EE} and R_C can be equal to, *or* larger than, r_o). The analysis is straightforward but tedious and can be shown (Problem 9.79) to yield the following result

$$R_{icm} \simeq \beta R_{EE} \frac{1 + R_C/\beta r_o}{1 + \dfrac{R_C + 2R_{EE}}{r_o}} \tag{9.96}$$

Example 9.5

For the differential amplifier analyzed in Example 9.3, let the bias-current source have an output resistance $R_{EE} = 200$ kΩ. Evaluate:

(a) the worst-case common-mode gain if the two collector resistances are accurate to within ±1%.
(b) the CMRR in dB.
(c) the input common-mode resistance (assuming the Early voltage $V_A = 100$ V).

Solution

First we observe that the two emitter resistances R_E will have negligible effect on A_{cm}.

(a) Using Eq. (9.94),

$$|A_{cm}| = \frac{R_C}{2R_{EE}} \frac{\Delta R_C}{R_C}$$

where $\Delta R_C = 0.02R_C$ in the worst case. Thus,

$$|A_{cm}| = \frac{10}{2 \times 200} \times 0.02 = 5 \times 10^{-4} \text{ V/V}$$

(b)

$$\text{CMRR} = 20 \log \frac{|A_d|}{|A_{cm}|}$$

where from Example 9.3, $|A_d| = 40$, thus

$$\text{CMRR} = 20 \log \frac{40}{5 \times 10^{-4}} = 98 \text{ dB}$$

(c)

$$r_o = \frac{V_A}{I/2} = \frac{100}{0.5} = 200 \text{ k}\Omega$$

Using Eq. (9.96),

$$R_{in} = 6.7 \text{ M}\Omega$$

EXERCISE

9.14 A bipolar differential amplifier uses a simple (i.e., a single CE transistor) current source to supply a bias current I of 200 μA, and simple current-source loads formed by *pnp* transistors. For all transistors, $\beta = 100$ and $|V_A| = 10$ V. Find g_m, R_C, $|A_d|$, R_{id}, R_{EE}, CMRR (if the two load transistors exhibit a 1% mismatch in their r_o's), and R_{icm}. (*Hint*: Remember to take into account r_{o1} and r_{o2} in calculating $|A_d|$.)

Ans. 4 mA/V; 100 kΩ; 200 V/V; 50 kΩ, 50 kΩ; 86 dB; 1.68 MΩ

9.4 DC Offset

Because differential amplifiers are directly coupled and have finite gain at dc, they suffer from a number of dc problems. In this section we study some of these.

9.4.1 Input Offset Voltage of the MOS Differential Amplifier

Consider the basic MOS differential amplifier in Fig. 9.27(a). If both input terminals are grounded, and if transistors Q_1 and Q_2 are perfectly matched, then the current I splits equally between Q_1 and Q_2, resulting in $I_{D1} = I_{D2} = I/2$. If the two load resistances R_{D1} and R_{D2} are exactly equal, then the dc voltages at the two drains will be equal, and the differential dc output voltage V_O will be zero.

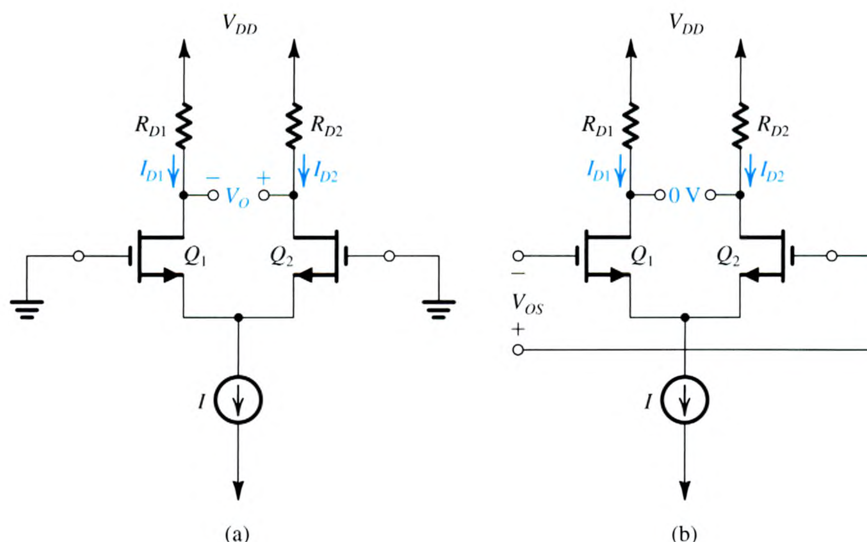


Figure 9.27 **(a)** The MOS differential pair with both inputs grounded. Device and resistance mismatches result in a finite dc output V_O. **(b)** Application of the input offset voltage V_{OS} to the input terminals with appropriate polarity reduces V_O to zero.

Practical circuits, however, exhibit mismatches between Q_1 and Q_2 that will cause a mismatch between I_{D1} and I_{D2}. As well, R_{D1} and R_{D2} will never be perfectly matched. These various mismatches, which are inevitable in a differential amplifier, result in a finite nonzero dc differential output voltage V_O even when the two input terminals are grounded. We call this voltage the **output dc offset voltage**. Since V_O is a result of random mismatches, we cannot deduce its polarity theoretically.

We will now consider the various possible mismatches in the differential amplifier circuit of Fig. 9.27(a) and estimate their contribution to the dc offset of the amplifier. We will do so by considering a parameter known as the **input dc offset voltage** V_{OS}, which is defined as the magnitude of the dc voltage, which, when applied between the two input terminals of the amplifier with appropriate polarity, will reduce the dc offset at the output, V_O, to zero. This definition is illustrated in Fig. 9.27(b). Since the polarity of V_O is not known a priori, we will not concern ourselves with the polarity of V_{OS}. Instead, our goal is to determine the magnitude of V_{OS} that results from each source of mismatch in the amplifier circuit.

Three factors contribute to the dc offset voltage of the MOS differential amplifier of Fig. 9.27: mismatch in the load resistances R_D, mismatch in the transconductance parameters k_n of Q_1 and Q_2 [recall that $k_n = \mu_n C_{ox}(W/L)$], and mismatch in the threshold voltages V_t of Q_1 and Q_2. Let's consider these three sources of offset in turn:

V_{OS} Due to R_D Mismatch

Let Q_1 and Q_2 be perfectly matched, so that $I_{D1} = I_{D2} = I/2$. Now, if the two resistances R_{D1} and R_{D2} have a mismatch ΔR_D, that is,

$$R_{D1} = R_D + \frac{\Delta R_D}{2}$$

$$R_{D2} = R_D - \frac{\Delta R_D}{2}$$

then a dc offset voltage will develop between the two drains. To reduce the output voltage to zero, we apply a voltage V_{OS} between the input terminals, as shown in Fig. 9.27(b). Thus the currents become

$$I_{D1} = \frac{I}{2} - g_m(V_{OS}/2)$$

$$I_{D2} = \frac{I}{2} + g_m(V_{OS}/2)$$

To obtain a zero voltage between the two output terminals, the voltage drops across R_{D1} and R_{D2} must be equal, that is,

$$\left(\frac{I}{2} - g_m\frac{V_{OS}}{2}\right)\left(R_D + \frac{\Delta R_D}{2}\right) = \left(\frac{I}{2} + g_m\frac{V_{OS}}{2}\right)\left(R_D - \frac{\Delta R_D}{2}\right)$$

which results in

$$V_{OS} = \left(\frac{I/2}{g_m}\right)\left(\frac{\Delta R_D}{R_D}\right)$$

Substituting for $g_m = \dfrac{2(I/2)}{V_{OV}}$ gives us

$$V_{OS} = \left(\frac{V_{OV}}{2}\right)\left(\frac{\Delta R_D}{R_D}\right) \tag{9.97}$$

Thus, the offset voltage is directly proportional to V_{OV} and, of course, to $\Delta R_D/R_D$. For example, a differential pair in which the two transistors are operating at an overdrive voltage of 0.2 V, and each drain resistance is accurate to within $\pm 1\%$, can have a worst case $\Delta R_D/R_D$ of 0.02 and correspondingly an input offset voltage V_{OS} of

$$|V_{OS}| = \left(\frac{V_{OV}}{2}\right)\left(\frac{\Delta R_D}{R_D}\right) = \frac{0.2}{2} \times 0.02 = 2\text{ mV}$$

V_{OS} Due to k_n Mismatch Next, consider the case when the transconductance parameters k_n of Q_1 and Q_2 exhibit a mismatch Δk_n, that is,

$$k_{n1} = k_n + \frac{1}{2}\Delta k_n$$

$$k_{n2} = k_n - \frac{1}{2}\Delta k_n$$

Such a mismatch can be caused, for example, by a mismatch in the (W/L) ratios of Q_1 and Q_2. The k_n mismatch will cause the current I to no longer divide equally between Q_1 and Q_2. Instead, with the two input terminals grounded and thus $V_{GS1} = V_{GS2}$, the currents become

$$I_{D1} = \frac{I}{2}\left(1 + \frac{1}{2}\frac{\Delta k_n}{k_n}\right)$$

$$I_{D2} = \frac{I}{2}\left(1 - \frac{1}{2}\frac{\Delta k_n}{k_n}\right)$$

The mismatch between I_{D1} and I_{D2} will result in an output offset voltage V_O even when the two resistances R_D are perfectly matched. To reduce V_O to zero, we apply an input voltage

V_{OS}, as shown in Fig. 9.27(b), and adjust the value of V_{OS} to reduce the mismatch between I_{D1} and I_{D2} to zero. This occurs when

$$g_m \frac{V_{OS}}{2} = \frac{I}{2} \times \frac{1}{2} \frac{\Delta k_n}{k_n}$$

which results in

$$V_{OS} = \left(\frac{I/2}{g_m}\right)\left(\frac{\Delta k_n}{k_n}\right)$$

Substituting for $g_m = \dfrac{2(I/2)}{V_{OV}}$ gives

$$V_{OS} = \left(\frac{V_{OV}}{2}\right)\left(\frac{\Delta k_n}{k_n}\right) \tag{9.98}$$

Here again, we note that V_{OS} resulting from a k_n mismatch is proportional to V_{OV} and, of course, $\Delta k_n / k_n$.

V_{OS} Due to V_t Mismatch Finally, consider the effect of a mismatch ΔV_t between the threshold voltages of Q_1 and Q_2, that is,

$$V_{t1} = V_t + \frac{\Delta V_t}{2}$$

$$V_{t2} = V_t - \frac{\Delta V_t}{2}$$

The current I_{D1} will be given by

$$I_{D1} = \frac{1}{2} k_n \left(V_{GS} - V_t - \frac{\Delta V_t}{2}\right)^2$$

$$= \frac{1}{2} k_n (V_{GS} - V_t)^2 \left(1 - \frac{\Delta V_t}{2(V_{GS} - V_t)}\right)^2$$

Substituting $V_{GS} - V_t = V_{OV}$ and assuming that $\Delta V_t \ll 2V_{OV}$, we can approximate the expression for I_{D1} as

$$I_{D1} \simeq \frac{1}{2} k_n V_{OV}^2 \left(1 - \frac{\Delta V_t}{V_{OV}}\right)$$

Now, substituting for $\frac{1}{2} k_n V_{OV}^2$ by $(I/2)$, we obtain

$$I_{D1} = \frac{I}{2}\left(1 - \frac{\Delta V_t}{V_{OV}}\right)$$

Since $I_{D2} = I - I_{D1}$, we obtain

$$I_{D2} = \frac{I}{2}\left(1 + \frac{\Delta V_t}{V_{OV}}\right)$$

To reduce the imbalance between I_{D1} and I_{D2} to zero and thus obtain a zero dc output offset voltage, we apply V_{OS} whose value is such that

$$g_m \frac{V_{OS}}{2} = \left(\frac{I}{2}\right)\left(\frac{\Delta V_t}{V_{OV}}\right)$$

That is,

$$V_{OS} = \frac{2(I/2)}{g_m V_{OV}} (\Delta V_t)$$

Substituting $g_m = \dfrac{2(I/2)}{V_{OV}}$ gives

$$V_{OS} = \Delta V_t \tag{9.99}$$

which is a very logical result! For modern CMOS technology, ΔV_t can be as high as a few millivolts.

Finally, since the three sources of offset voltage are *not* correlated, an estimate of the total input offset voltage can be found as

$$V_{OS} = \sqrt{\left(\frac{V_{OV}}{2}\frac{\Delta R_D}{R_D}\right)^2 + \left(\frac{V_{OV}}{2}\frac{\Delta k_n}{k_n}\right)^2 + (\Delta V_t)^2} \tag{9.100}$$

EXERCISE

9.15 For a MOS differential pair operated at $V_{OV} = 0.2$ V, find the three components of the input offset voltage. Let $\Delta R_D/R_D = 2\%$, $\Delta k_n/k_n = 2\%$, and $\Delta V_t = 2$ mV. Use Eq. (9.100) to obtain an estimate of the total V_{OS}.
Ans. 2 mV; 2 mV; 2 mV; 3.5 mV

9.4.2 Input Offset Voltage of the Bipolar Differential Amplifier

We can find the offset voltage of the bipolar differential pair in Fig. 9.28(a) the same way we determined the dc offset voltage of the MOS pair above. Note, however, that in the bipolar case there is no analog to the V_t mismatch of the MOSFET pair. Here the output offset results from mismatches in the load resistances R_{C1} and R_{C2} and from junction area, β, and other mismatches in Q_1 and Q_2. Consider first the effect of the load mismatch. Let

$$R_{C1} = R_C + \frac{\Delta R_C}{2}$$

$$R_{C2} = R_C - \frac{\Delta R_C}{2}$$

and assume that Q_1 and Q_2 are perfectly matched. It follows that current I will divide equally between Q_1 and Q_2, and thus a dc offset voltage will develop between the two collectors.

To reduce this output offset voltage to zero, we apply a voltage V_{OS} between the input terminals, as shown is Fig. 9.28(b). If the value of V_{OS} is such that the imbalance it introduces between I_{C1} and I_{C2} cancels the output dc offset due to ΔR_C, a development similar to that used above for the MOS case results in

$$V_{OS} = V_T\left(\frac{\Delta R_C}{R_C}\right) \tag{9.101}$$

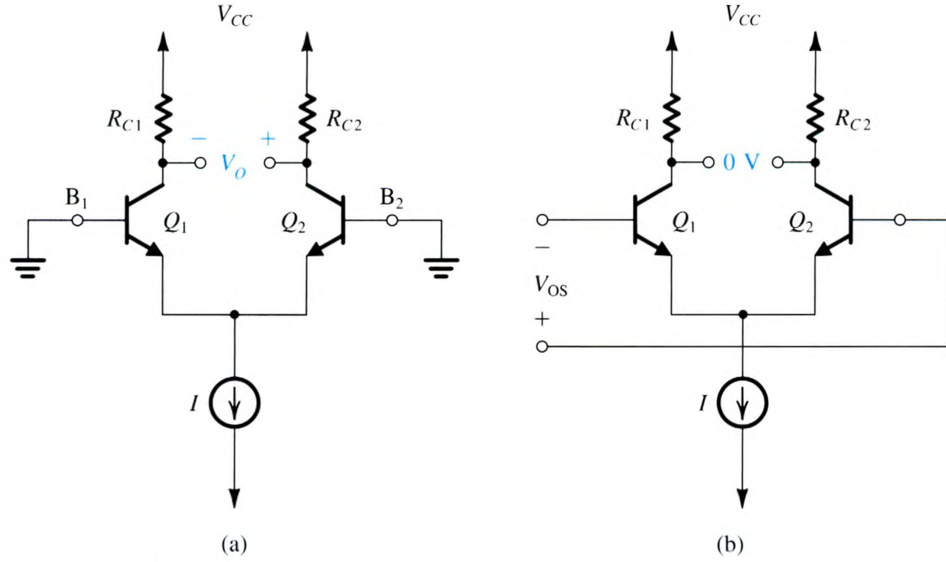


Figure 9.28 **(a)** The BJT differential pair with both inputs grounded. Device and resistance mismatch result in a finite dc output V_O. **(b)** Application of the input offset voltage V_{OS} to the input terminals with appropriate polarity reduces V_O to zero.

An important point to note is that in comparison to the corresponding expression for the MOS pair (Eq. 9.97), here the offset is proportional to V_T rather than $V_{OV}/2$. V_T at 25 mV is 3 to 6 times lower than $V_{OV}/2$. Hence bipolar differential pairs exhibit lower offsets than their MOS counterparts. As an example, consider collector resistors that are accurate to within $\pm 1\%$. Then the worst-case mismatch will be $\Delta R_C/R_C = 0.02$, and the resulting input offset voltage will be

$$V_{OS} = 25 \times 0.02 = 0.5 \text{ mV}$$

Next consider the effect of mismatches in transistors Q_1 and Q_2, specifically in their emitter–base junction areas. This kind of area mismatch gives rise to a proportional mismatch in the scale currents I_S,

$$I_{S1} = I_S + \frac{\Delta I_S}{2}$$

$$I_{S2} = I_S - \frac{\Delta I_S}{2}$$

Refer to Fig. 9.28(a) and note that $V_{BE1} = V_{BE2}$. Thus, the current I will split between Q_1 and Q_2 in proportion to their I_S values, resulting in

$$I_{E1} = \frac{I}{2}\left(1 + \frac{\Delta I_S}{2I_S}\right)$$

$$I_{E2} = \frac{I}{2}\left(1 - \frac{\Delta I_S}{2I_S}\right)$$

The corresponding mismatch between I_{C1} and I_{C2} will result in an output offset voltage V_O, even when the two resistances R_C are perfectly matched. To reduce V_O to zero, we apply

an input voltage V_{OS}, as shown in Fig. 9.28(b), and adjust its value to reduce the mismatch between I_{C1} and I_{C2} to zero. We can show that the value of V_{OS} is given by

$$V_{OS} = V_T\left(\frac{\Delta I_S}{I_S}\right) \tag{9.102}$$

As an example, an area mismatch of 4% gives rise to $\Delta I_S/I_S = 0.04$ and an input offset voltage of 1 mV. Here again we note that the offset voltage is proportional to V_T rather than to the much larger V_{OV}, which determines the offset of the MOS pair due to $\Delta k_n/k_n$ mismatch.

Since the two contributions to the input offset voltage are usually *not* correlated, we can estimate the total input offset voltage using

$$\begin{aligned} V_{OS} &= \sqrt{\left(V_T\frac{\Delta R_C}{R_C}\right)^2 + \left(V_T\frac{\Delta I_S}{I_S}\right)^2} \\ &= V_T\sqrt{\left(\frac{\Delta R_C}{R_C}\right)^2 + \left(\frac{\Delta I_S}{I_S}\right)^2} \end{aligned} \tag{9.103}$$

There are other possible sources for input offset voltage such as mismatches in the values of β and r_o. Some of these are investigated in the end-of-chapter problems. Finally, note that there is a popular way to compensate for the offset voltage. It involves introducing a deliberate mismatch in the values of the two collector resistances so that the differential output voltage is reduced to zero when both input terminals are grounded. Such an **offset-nulling** scheme is explored in Problem 9.81.

9.4.3 Input Bias and Offset Currents of the Bipolar Differential Amplifier

In a perfectly symmetric bipolar differential pair, the two input terminals carry equal dc currents; that is,

$$I_{B1} = I_{B2} = \frac{I/2}{\beta + 1} \tag{9.104}$$

This is the **input bias current** of the differential amplifier.

Mismatches in the amplifier circuit and most importantly a mismatch in β make the two input dc currents unequal. The resulting difference is the **input offset current**, I_{OS}, given as

$$I_{OS} = |I_{B1} - I_{B2}| \tag{9.105}$$

Let

$$\beta_1 = \beta + \frac{\Delta\beta}{2}$$
$$\beta_2 = \beta - \frac{\Delta\beta}{2}$$

then

$$I_{B1} = \frac{I}{2}\frac{1}{\beta + 1 + \Delta\beta/2} \simeq \frac{I}{2}\frac{1}{\beta + 1}\left(1 - \frac{\Delta\beta}{2\beta}\right)$$

$$I_{B2} = \frac{I}{2}\frac{1}{\beta + 1 - \Delta\beta/2} \simeq \frac{I}{2}\frac{1}{\beta + 1}\left(1 + \frac{\Delta\beta}{2\beta}\right)$$

$$I_{OS} = \frac{I}{2(\beta + 1)}\left(\frac{\Delta\beta}{\beta}\right) \tag{9.106}$$

Formally, the input bias current I_B is defined as follows:

$$I_B \equiv \frac{I_{B1} + I_{B2}}{2} = \frac{I}{2(\beta + 1)} \tag{9.107}$$

Thus

$$I_{OS} = I_B\left(\frac{\Delta\beta}{\beta}\right) \tag{9.108}$$

As an example, a 10% β mismatch results in an offset current that is one-tenth the value of the input bias current.

Finally, note that a great advantage of the MOS differential pair is that it does not suffer from a finite input bias current or from mismatches thereof!

EXERCISE

9.16 For a BJT differential amplifier utilizing transistors having $\beta = 100$, matched to 10% or better, and areas that are matched to 10% or better, along with collector resistors that are matched to 2% or better, find V_{OS}, I_B, and I_{OS}. The dc bias current I is 100 μA.

Ans. 2.55 mV; 0.5 μA; 50 nA

9.4.4 A Concluding Remark

The definitions presented here are identical to those presented in Chapter 2 for op amps. In fact, as you will see in Chapter 13, it is the input differential stage in an op-amp circuit that primarily determines the op-amp dc offset voltage, input bias and offset currents, and input common-mode range.

9.5 The Differential Amplifier with a Current-Mirror Load

The differential amplifiers we have studied thus far have been of the differential output variety; that is, the output is taken between the two drains (or two collectors) rather than between one of the drains (collectors) and ground. Taking the output differentially has three major advantages:

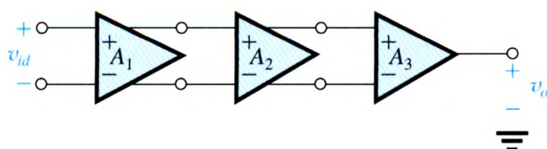


Figure 9.29 A three-stage amplifier consisting of two differential-in, differential-out stages, A_1 and A_2, and a differential-in, single-ended-out stage A_3.

1. It decreases the common-mode gain and thus increases the common-mode rejection ratio (CMRR). Recall that while the drain (collector) voltages change somewhat in response to a common-mode input signal, the difference between the drain (collector) voltages remains essentially zero except for a small change due to the mismatches inevitably present in the circuit.
2. It decreases the input offset voltage.
3. It increases the differential gain by a factor of 2 (6 dB) because the output is the difference between two voltages of equal magnitude and opposite sign.

These advantages are sufficiently compelling that at least the first stage in an IC amplifier such as an op amp is **differential-in, differential-out**. The differential transmission of the signal on the chip also minimizes its susceptibility to corruption with noise and interference, which usually occur in a common-mode fashion. Nevertheless, it is usually required at some point to convert the signal from differential to single-ended; for instance, to connect it to an off-chip load. Figure 9.29 shows a block diagram of a three-stage amplifier in which the first two stages are of the differential-in, differential-out type, and the third has a single-ended output, that is, an output that is referenced to ground. We now address the question of conversion from differential to single-ended.

9.5.1 Differential-to-Single-Ended Conversion

Figure 9.30 illustrates the simplest, most basic approach for differential-to-single-ended conversion. It consists of simply ignoring the drain current signal of Q_1 and eliminating its drain resistor altogether, and taking the output between the drain of Q_2 and ground. The obvious drawback of this scheme[1] is that we lose a factor of 2 (or 6 dB) in gain as a result of "wasting" the drain signal current of Q_1. A much better approach would be to find a way of utilizing the drain current signal of Q_1, and that is exactly what the circuit we are about to discuss accomplishes.

9.5.2 The Current-Mirror-Loaded MOS Differential Pair

Figure 9.31(a) shows a MOS differential pair formed by transistors Q_1 and Q_2, loaded by a current mirror formed by transistors Q_3 and Q_4. As shown, the output voltage is taken between the drain connection of Q_2 and Q_4 and ground.

To see how the circuit in Fig. 9.31(a) works, we first consider an idealized situation with perfect matching and with the channel-length modulation (and hence r_o) in all transistors neglected. Figure 9.31(b) shows the quiescent or equilibrium situation, where the two input terminals are connected to ground. The bias current I divides equally between Q_1 and Q_2.

[1]The circuit in Fig. 9.30 would be totally unacceptable as the input stage of a differential amplifier such as an op amp. This is because its CMRR is very low!

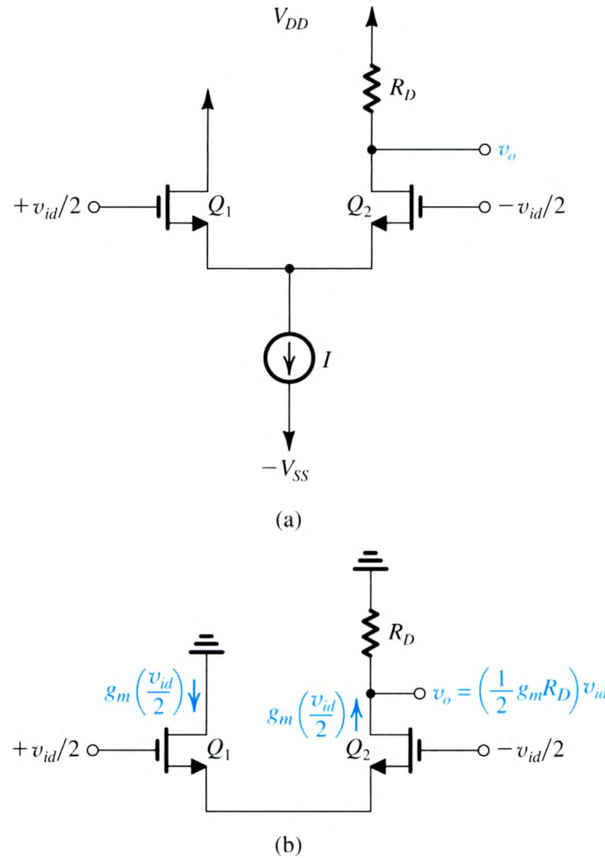


Figure 9.30 **(a)** A simple but inefficient approach for differential-to-single-ended conversion. **(b)** The circuit with the power supplies eliminated for signal calculations. Note that the current signal in Q_1 is not utilized, and as a result, the gain achieved is 6 dB lower than the differential-output case.

The drain current of Q_1, $I/2$, is fed to the input transistor of the mirror, Q_3. Thus, a replica of this current is provided by the output transistor of the mirror, Q_4. Notice that at the output node, the two currents $I/2$ balance each other out, leaving a zero current to flow out to the next stage or to a load (not shown). This is obviously the desired result! Further, if Q_4 is perfectly matched to Q_3, its drain voltage will track the voltage at the drain of Q_3; thus, in equilibrium, the voltage at the output will be $V_O = V_{DD} - V_{SG3}$.

Next, consider the situation in Figure 9.31(c), where a common-mode signal v_{icm} is applied to the input terminals. Here, we have eliminated all dc sources and replaced the bias current source I by its output resistance R_{SS}. As you can see, in response to v_{icm}, Q_1 and Q_2 conduct current signals $i \simeq v_{icm}/2R_{SS}$. The drain current signal of Q_1 is reproduced by the current mirror in the drain of Q_4. At the output mode, the mirrored current i cancels out the signal current i of Q_2, resulting in a zero output current and correspondingly zero output voltage. Thus, the amplifier rejects the common-mode input signal v_{icm}.

Finally, Fig. 9.31(d) shows the situation when a differential input signal v_{id} is applied. As expected, a current $i = v_{id}/(\frac{2}{g_m})$ develops in Q_1 and Q_2. The drain current signal of Q_1 is reproduced by the current mirror in the drain of Q_4. In this case, however, at the output node, the mirrored current i *adds* to the current signal i in the drain of Q_2, thus providing

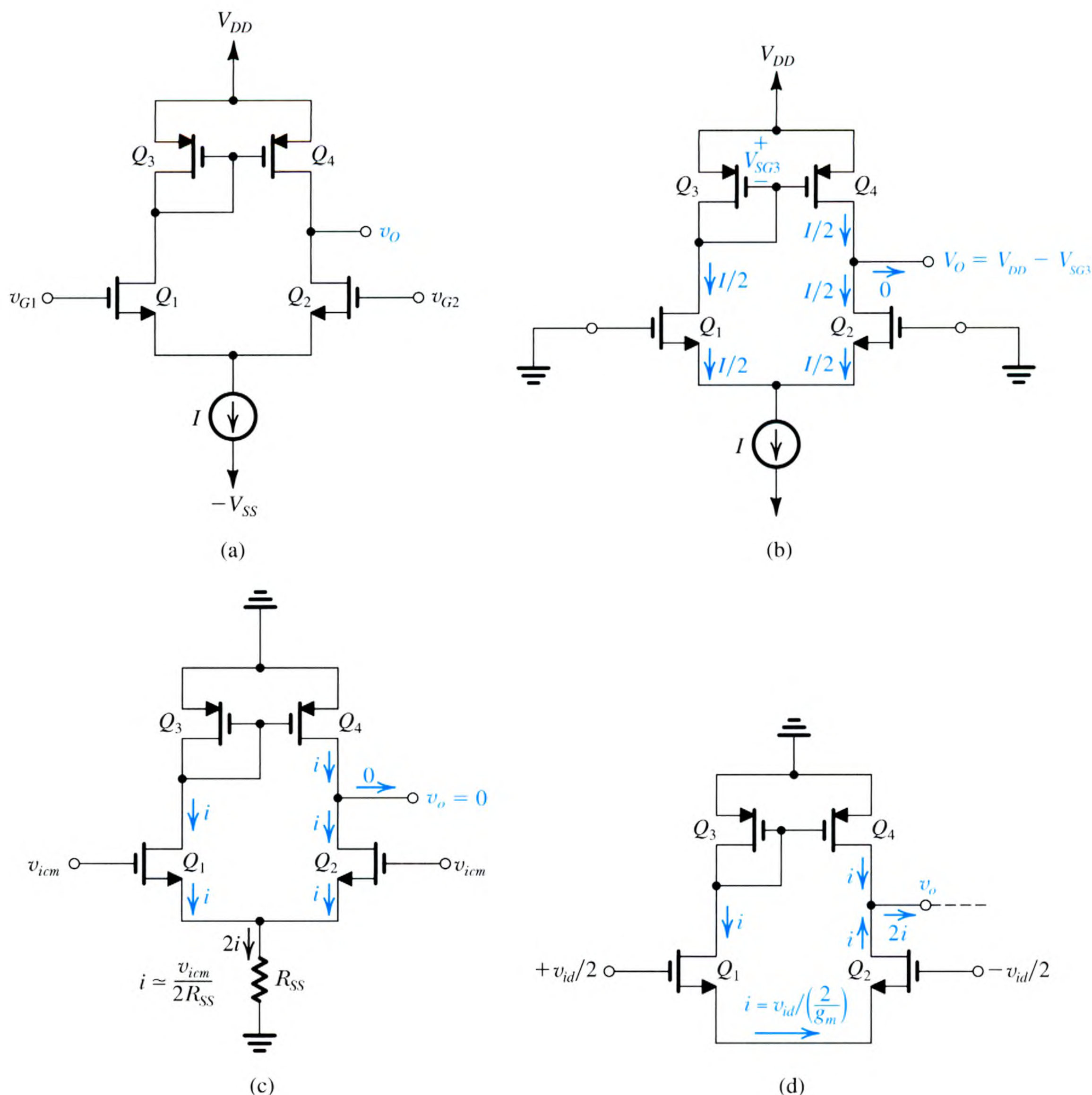


Figure 9.31 **(a)** The current-mirror-loaded MOS differential pair. Assuming perfect matching and neglecting the Early effect, in all transistors, we show in **(b)** the circuit at equilibrium, **(c)** the circuit with a common-mode input signal v_{icm}, and **(d)** the circuit with a differential input signal v_{id}.

an output current $2i$. It is this factor of 2, which is a result of the current-mirror action, that makes it possible to convert the signal to single-ended (that is, between the output node and ground) with no loss of gain! The output current $2i$ flows through the parallel combination of the output resistance of the amplifier and a load resistance (if one is connected) to provide the output voltage v_o.

In conclusion, before immersing ourselves in detailed analysis of the circuit, it is important to recap the essence of its operation: For dc quantities and common-mode inputs,

the current-mirror load produces an output current in the drain of Q_4 that *cancels* the current of Q_2. On the other hand, for differential input signals, the output current of the mirror *adds* to the current of Q_2.

9.5.3 Differential Gain of the Current-Mirror-Loaded MOS Pair

As we learned in Chapter 8, the Early effect manifested by the output resistance r_o of the transistor, plays a significant role in the operation of active-loaded amplifiers. Therefore, we shall now take r_o into account and derive an expression for the differential gain v_o/v_{id} of the current-mirror-loaded MOS differential pair (Fig. 9.31a). To that end, we first observe that the circuit is *not* symmetrical: While the drain of Q_1 sees the small resistance of the diode-connected transistor Q_3 (approximately equal to $1/g_{m3}$), the drain of Q_2 sees the much larger output resistance of Q_4 (r_{o4}). Thus, a virtual ground will *not* develop at the common sources[2] and we cannot use the differential half-circuit technique.

Our approach will be to represent the output of the circuit in Fig. 9.31(a) by the general equivalent circuit shown in Fig. 9.32. Here G_{md} is the short-circuit transconductance for differential input, and R_o is the output resistance. In the following, we will show that

$$G_{md} = g_{m1,2} \tag{9.109}$$

where $g_{m1,2}$ is the transconductance of each of Q_1 and Q_2. We will also show that

$$R_o = r_{o2} \,\|\, r_{o4} \tag{9.110}$$

Since r_{o2} is the output resistance of the differential pair and r_{o4} is the output resistance of the current mirror, R_o can be expressed as

$$R_o = R_{od} \,\|\, R_{om} \tag{9.110A}$$

Thus we will have two intuitively appealing results: The differential short-circuit transconductance of the circuit is equal to g_m of each of the two transistors of the differential pair, and the output resistance is the parallel equivalent of the output resistance of the differential amplifier R_{od} and the output resistance of the current mirror R_{om}. Thus, the open-circuit differential voltage gain can be found as

$$A_d \equiv \frac{v_o}{v_{id}} = G_{md}R_o = g_{m1,2}(r_{o2} \,\|\, r_{o4}) \tag{9.111}$$

Alternatively we can express A_d as

$$A_d = g_{m1,2}(R_{od} \,\|\, R_{om}) \tag{9.111A}$$

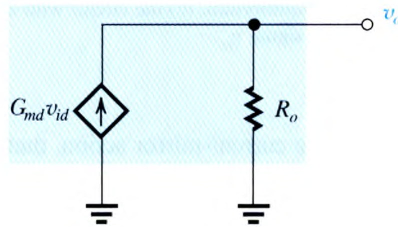


Figure 9.32 Output equivalent circuit of the amplifier in Fig. 9.31(a) for differential input signals.

[2]The qualitative description of circuit operation above implied that a virtual ground develops at the MOSFET sources. That was the case because we were neglecting r_o of all transistors.

Writing $g_{m1,2}$ simply as g_m, and for the case $r_{o2} = r_{o4} = r_o$,

$$A_d = \frac{1}{2} g_m r_o = \frac{1}{2} A_0 \quad (9.112)$$

where A_0 is the intrinsic gain of the MOS transistor.

EXERCISE

9.17 A current-mirror-loaded MOS differential amplifier of the type shown in Fig. 9.31(a) is specified as follows: $(W/L)_n = 100$, $(W/L)_p = 200$, $\mu_n C_{ox} = 2\mu_p C_{ox} = 0.2$ mA/V^2, $V_{An} = |V_{Ap}| = 20$ V, and $I = 0.8$ mA. Calculate G_{md}, R_o, and A_d.

Ans. 4 mA/V; 25 kΩ; 100 V/V

Derivation of the Differential Short-Circuit Transconductance, G_{md} Figure 9.33(a) shows the current-mirror-loaded MOS amplifier with the output terminal short-circuited to ground. Our purpose is to determine the differential short-circuit transconductance

$$G_{md} \equiv \frac{i_o}{v_{id}}$$

Short-circuiting the output terminal makes the circuit nearly balanced. This is because the drain of Q_1 sees the small resistance of the diode-connected transistor Q_3, and now the drain of Q_2 sees a short circuit. It follows that the voltage at the MOSFET sources will be approximately zero. Now, replacing each of the four transistors with its hybrid-π model and noting that for the diode-connected transistor Q_3, the model reduces to a resistance $(1/g_{m3} \| r_{o3})$, we obtain the equivalent circuit shown in Fig. 9.33(b). The short-circuit output current i_o can be found by writing a node equation at the output and noting that the currents in r_{o2} and r_{o4} are zero; thus

$$i_o = g_{m2}\left(\frac{v_{id}}{2}\right) - g_{m4}\, v_{gs4} \quad (9.113)$$

Next, we note that

$$v_{gs4} = v_{gs3} \quad (9.114)$$

and v_{gs3} can be obtained from a node equation at d_1 as

$$v_{gs3} = -g_{m1}\left(\frac{v_{id}}{2}\right)\left(\frac{1}{g_{m3}} \| r_{o3} \| r_{o1}\right)$$

which for the usual case of $\frac{1}{g_{m3}} \ll r_{o3}, r_{o1}$, reduces to

$$v_{gs3} \simeq -\frac{g_{m1}}{g_{m3}}\left(\frac{v_{id}}{2}\right) \quad (9.115)$$

Combining Eqs. (9.113) to (9.115) and substituting $g_{m3} = g_{m4}$ and $g_{m1} = g_{m2} = g_m$ gives

$$i_o = g_m v_{id}$$

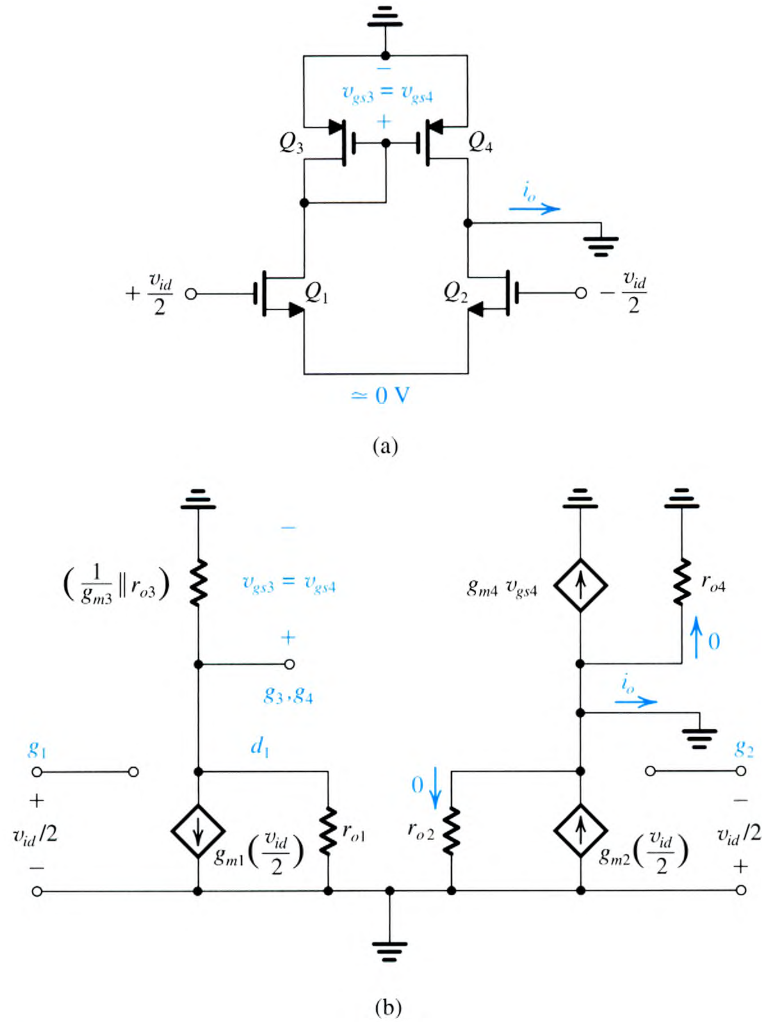


Figure 9.33 Derivation of the differential short-circuit transconductance $G_{md} \equiv i_o/v_{id}$.

from which G_{md} is found to be

$$G_{md} = g_m$$

as expected.

Derivation of the Output Resistance R_o Figure 9.34 shows the circuit[3] for determining the output resistance R_o. Observe that we have set v_{id} to zero, resulting in the ground connections at the gates of Q_1 and Q_2. We have applied a test voltage v_x in order

[3]Note that rather than replacing each transistor with its small-signal model, we are, for simplicity, using the models implicitly. Thus we have "pulled r_o out" of each transistor and shown it separately so that the drain current becomes $g_m v_{gs}$.

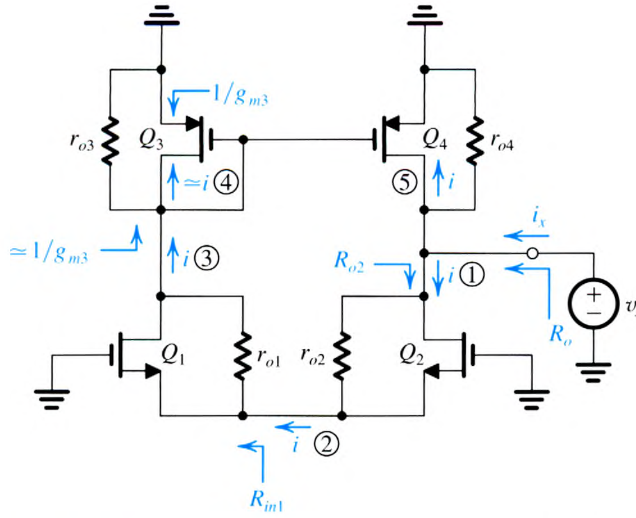


Figure 9.34 Circuit for determining R_o. The circled numbers indicate the order of the analysis steps.

to determine R_o,

$$R_o \equiv \frac{v_x}{i_x}$$

We can greatly simplify our analysis by following the circled numbers to observe the current transmission around the circuit. The current i that enters Q_2 must exit at its source. It then enters Q_1, exiting at the drain to feed the Q_3–Q_4 mirror. Since for the diode-connected transistor Q_3, $1/g_{m3}$ is much smaller than r_{o3}, most of the current i flows into the drain proper of Q_3. The mirror responds by providing an equal current i in the drain of Q_4.

We can find the relationship between i and v_x by observing that at the output node

$$i = v_x/R_{o2}$$

where R_{o2} is the output resistance of Q_2. Now, Q_2 is a CG transistor and has in its source lead the input resistance R_{in1} of the CG transistor Q_1. Noting that the load resistance of Q_1 is $[(1/g_{m3}) \| r_{o3}]$, which is approximately $1/g_{m3}$, we can obtain R_{in1} by using the expression for the input resistance of a CG transistor (Eq. 8.52),

$$R_{\text{in1}} = \frac{r_{o1} + R_L}{g_{m1} r_{o1}} = \frac{r_{o1} + (1/g_{m3})}{g_{m1} r_{o1}}$$
$$= \frac{1}{g_{m1}} + \frac{1/g_{m3}}{g_{m1} r_{o1}} \simeq \frac{1}{g_{m1}}$$

We then use this value of R_{in1} to determine R_{o2} utilizing the expression in Eq. (8.55) as follows:

$$R_{o2} = R_{\text{in1}} + r_{o2} + g_{m2} r_{o2} R_{\text{in1}}$$
$$= \frac{1}{g_{m1}} + r_{o2} + \left(\frac{g_{m2}}{g_{m1}}\right) r_{o2}$$

which, for $g_{m1} = g_{m2} = g_m$ and $g_{m2} r_{o2} \gg 1$, yields

$$R_{o2} \simeq 2r_{o2} \tag{9.116}$$

Returning to the output node, we write

$$i_x = i + i + \frac{v_x}{r_{o4}}$$
$$= 2i + \frac{v_x}{r_{o4}} = 2\frac{v_x}{R_{o2}} + \frac{v_x}{r_{o4}}$$

Substituting for R_{o2} from Eq. (9.116), we obtain

$$i_x = 2\frac{v_x}{2r_{o2}} + \frac{v_x}{r_{o4}}$$

Thus,

$$R_o \equiv \frac{v_x}{i_x} = r_{o2} \parallel r_{o4} \tag{9.117}$$

which is the result we stated earlier.

9.5.4 The Bipolar Differential Pair with a Current-Mirror Load

The bipolar version of the active-loaded differential pair is shown in Fig. 9.35(a). The circuit structure and operation are very similar to those of its MOS counterpart except that here we have to contend with the effects of finite β and the resulting finite input resistance at the base, r_π. For the time being, however, we shall ignore the effect of finite β on the dc bias of the four transistors and assume that in equilibrium all transistors are operating at a dc current of $I/2$.

Differential Gain To obtain an expression for the differential gain, we use an approach identical to the one we used earlier in the MOS case. That is, we represent the output of the amplifier with the equivalent circuit shown in Fig. 9.35(b) and show that the differential short-circuit transconductance G_{md} is given by

$$G_{md} = g_{m1,2} \tag{9.118}$$

where $g_{m1,2}$ denotes g_m of each of Q_1 and Q_2, and that the output resistance R_o is given by

$$R_o = r_{o2} \parallel r_{o4} = R_{od} \parallel R_{om} \tag{9.119}$$

Both results are identical to those for the MOS case and can be similarly derived.

Equations (9.118) and (9.119) can now be combined to obtain the differential gain,

$$A_d \equiv \frac{v_o}{v_{id}} = G_{md}R_o = g_m(r_{o2} \parallel r_{o4}) \tag{9.120}$$

where $g_m = g_{m1} = g_{m2} \simeq \dfrac{I/2}{V_T}$, and since $r_{o2} = r_{o4} = r_o$, we can simplify Eq. (9.120) to

$$A_d = \frac{1}{2}g_m r_o \tag{9.121}$$

Although this expression is identical to the one we found for the MOS circuit, the gain is much larger because $g_m r_o$ for the BJT is more than an order of magnitude greater than $g_m r_o$ of a MOSFET. The downside, however, lies in the low input resistance of BJT amplifiers. Indeed, from the circuit in Fig. 9.35(a), we can see that the differential input resistance is equal to $2r_\pi$,

$$R_{id} = 2r_\pi \tag{9.122}$$

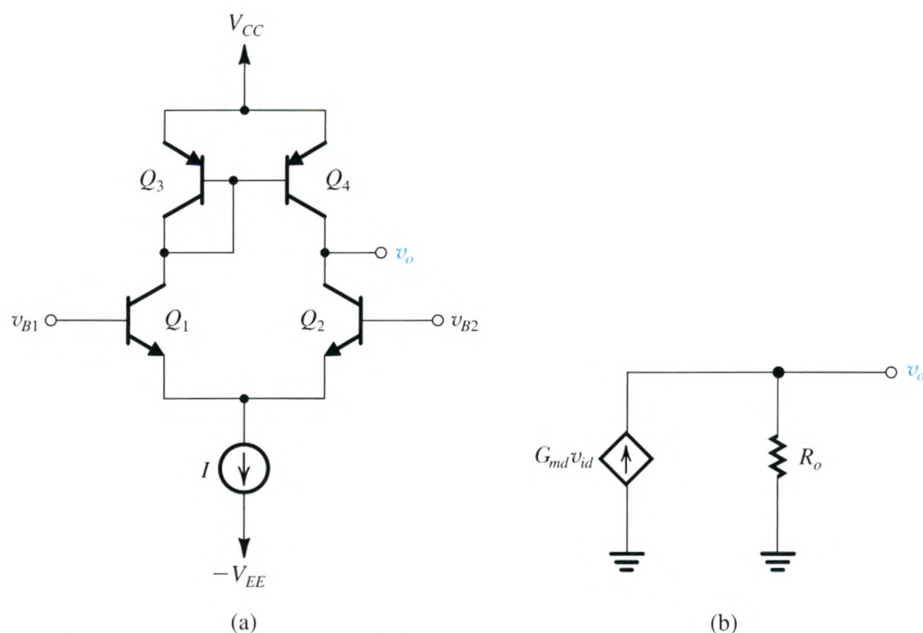


Figure 9.35 **(a)** Current-mirror-loaded bipolar differential pair. **(b)** Small-signal equivalent circuit of the amplifier output when a differential signal $v_{id} \equiv v_{B1} - v_{B2}$ is applied.

in sharp contrast to the infinite input resistance of the MOS amplifier. Thus, while the voltage gain realized in a current-mirror-loaded BJT amplifier stage is large, when a subsequent BJT stage is connected to the output, its inevitably low input resistance will drastically reduce the overall voltage gain.

EXERCISE

9.18 For the current-mirror-loaded BJT differential amplifier let $I = 0.8$ mA, $V_A = 100$ V, and $\beta = 160$. Find G_{md}, R_o, A_d, and R_{id}.
Ans. 16 mA/V; 125 kΩ; 2000 V/V; 20 kΩ

Systematic Input Offset Voltage In addition to the random offset voltages that result from the mismatches inevitably present in the differential amplifier, the current-mirror-loaded bipolar differential pair suffers from a systematic offset voltage that has no counterpart in the MOS version. This is due to the error in the current transfer ratio of the current-mirror load caused by the finite β of the *pnp* transistors that make up the mirror. This issue is investigated in Problem 9.97.

9.5.5 Common-Mode Gain and CMRR

In our initial discussion of the current-mirror-loaded differential amplifier we found the common-mode gain to be zero. Specifically, Fig. 9.31(c) shows that the current mirror provides an output current in the drain of Q_4 that is exactly equal to the common-mode current signal in the drain of Q_2, resulting in a zero output current and correspondingly zero output voltage. This perfect cancellation of currents, however, is predicated on two assumptions: (1) that the common-mode current signals in the drains of Q_1 and Q_2 are equal, and (2) that the current transfer ratio of the current mirror is exactly one. The first assumption is not valid because of the obvious imbalance between the two sides of the circuit. The second assumption also is not valid for practical current-mirror circuits, as we have already seen in Section 8.2.5, where we derived an expression for the small-signal current transfer ratio of the basic MOS mirror and found that it deviates somewhat from unity. It follows that the common-mode gain of the current-mirror-loaded differential amplifier will not be zero. In the following we derive expressions for A_{cm} and for the CMRR.

Figure 9.36(a) shows the circuit with a common-mode signal v_{icm} applied and with the dc sources eliminated, except, of course, for the output resistance R_{SS} of the bias current source I. Because of the imbalance between the two sides of the circuit, we will not be able to use the common-mode half-circuit. Rather, as we have done in our analysis of the differential gain, we will model the common-mode operation of the circuit with the output equivalent circuit shown in Fig. 9.36(b). Here G_{mcm} is the *common-mode short-circuit transconductance*, and R_o is the output resistance. Of course, R_o is the same output resistance we derived in connection with the analysis of the differential gain above.

Derivation of the Common-Mode Transconductance G_{mcm} We can derive an expression for the common-mode transconductance G_{mcm} using the equivalent circuit in Fig. 9.36(c). Here we have replaced each of Q_1 and Q_2 with their equivalent-circuit models, and replaced the current mirror with its small-signal model. The latter consists of the input resistance R_{im}, the short-circuit current gain A_m, and the output resistance R_{om}. To obtain G_{mcm} we have short-circuited the output node to ground, thus

$$G_{mcm} \equiv \frac{i_o}{v_{icm}}$$

We analyze the circuit in Fig. 9.36(c) as follows. A node equation at D_1 yields

$$i_1 = g_{m1}\,(v_{icm} - v_s) + \frac{-i_1 R_{im} - v_s}{r_{o1}} \tag{9.123}$$

A node equation at D_2 provides

$$i_2 = g_{m2}\,(v_{icm} - v_s) - \frac{v_s}{r_{o2}} \tag{9.124}$$

Finally, a node equation at S results in

$$v_s = (i_1 + i_2)R_s \tag{9.125}$$

Next we use Eq. (9.125) to substitute for v_s into Eqs. (9.123) and (9.124). Then we solve the two resulting equations together to obtain i_1 and i_2. This process, which is rather tedious, yields the following results:

$$i_1 = \left(\frac{v_{icm}}{2R_{SS}}\right)\left(\frac{1}{D}\right) \tag{9.126}$$

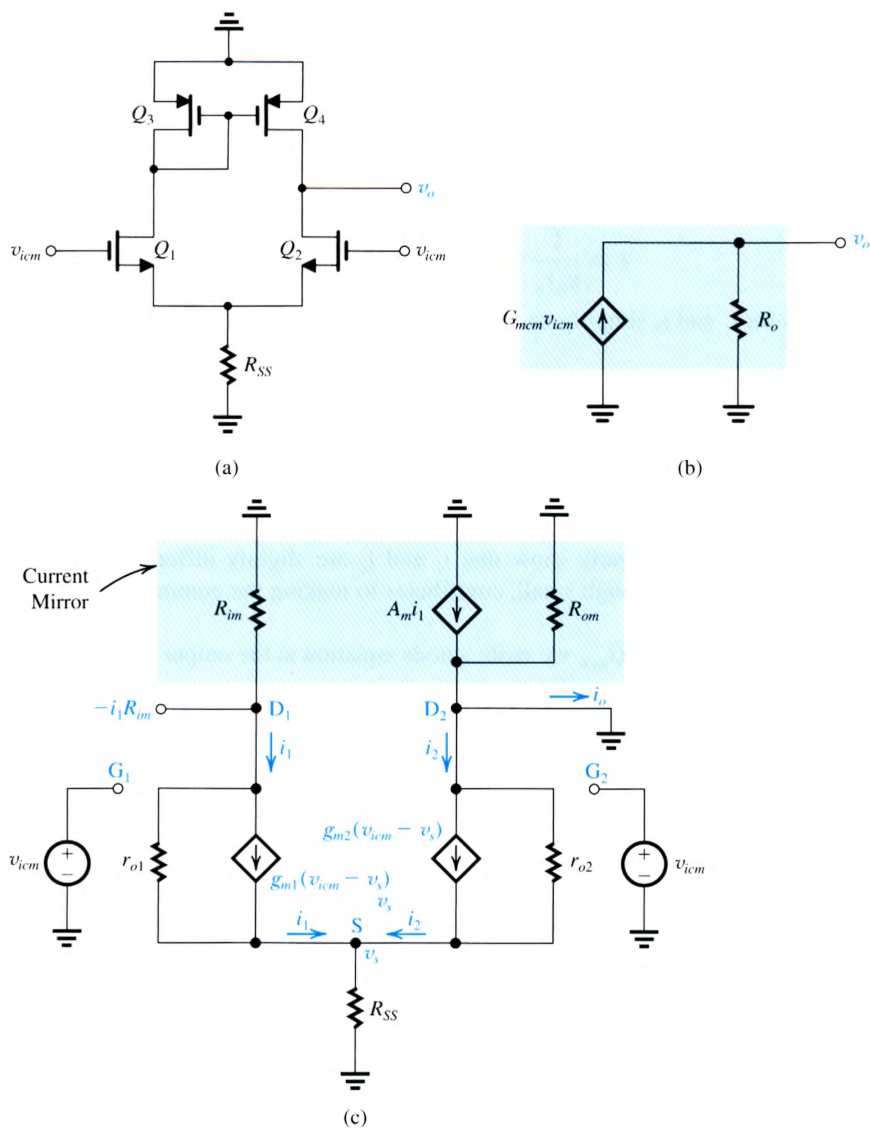


Figure 9.36 Analysis of the current-mirror-loaded MOS differential amplifier to determine its common-mode gain. **(a)** The circuit with v_{icm} applied and the dc sources removed. **(b)** Output equivalent circuit. **(c)** Circuit for determining G_{mcm}.

and

$$i_2 = \left(\frac{v_{icm}}{2R_{SS}}\right)\left(\frac{1+\dfrac{R_{im}}{r_o}}{D}\right) \tag{9.127}$$

where

$$D = 1 + \frac{1}{g_m r_o} + \frac{1}{2g_m R_{SS}} + \left(\frac{R_{im}}{2r_o}\right)\left(1 + \frac{1}{g_m r_o} + \frac{1}{g_m R_{SS}}\right) \tag{9.128}$$

In obtaining Eqs. (9.126), (9.127), and (9.128) we assumed $g_{m1} = g_{m2} = g_m$ and $r_{o1} = r_{o2} = r_o$. Since $g_m r_o \gg 1$, $g_m R_{SS} \gg 1$, and $R_{im} \ll r_o$,

$$D \simeq 1 + \epsilon + \frac{R_{im}}{2r_o}$$

where

$$\epsilon = \frac{1}{g_m r_o} + \frac{1}{2g_m R_{SS}} \ll 1$$

and the expressions for i_1 and i_2 simplify to

$$i_1 \simeq \frac{v_{icm}}{2R_{SS}}\left(1 - \epsilon - \frac{R_{im}}{2r_o}\right) \tag{9.129}$$

$$i_2 \simeq \left(\frac{v_{icm}}{2R_{SS}}\right)\left(1 - \epsilon + \frac{R_{im}}{2r_o}\right) \tag{9.130}$$

Equations (9.129) and (9.130) clearly show that i_1 and i_2 are slightly different [note that $(R_{im}/r_o) \ll 1$]. This difference, though small, contributes to making the common-mode gain nonzero.

To complete the derivation of G_{mcm} we write a node equation at the output node as

$$i_o = A_m i_1 - i_2 \tag{9.131}$$

Substituting for i_1 and i_2 from (9.129) and (9.130), respectively, gives

$$i_o = \frac{v_{icm}}{2R_{SS}}\left[(A_m - 1)(1 - \epsilon) - \frac{R_{im}}{2r_o}(A_m + 1)\right] \tag{9.132}$$

Thus,

$$G_{mcm} \equiv \frac{i_o}{v_{icm}} = \frac{1}{2R_{SS}}\left[(A_m - 1)(1 - \epsilon) - \frac{R_{im}}{2r_o}(A_m + 1)\right] \tag{9.133}$$

Equation (9.133) tells us the two separate reasons for the finite nonzero G_{mcm}: that $A_m \neq 1$ (the first term between the square brackets) and that i_1 and i_2 are unequal (the second term within the square brackets).

The expression in Eq. (9.133) is general and applies regardless of the actual current-mirror implementation used. For the simple MOS mirror used in the circuit in Fig. 9.36(a), we have from Section 8.5.2

$$R_{im} = \frac{1}{g_{m3}} \parallel r_{o3} = \frac{1}{g_{m3} + \frac{1}{r_{o3}}} \tag{9.134}$$

and

$$A_m = 1 \Big/ \left[1 + \frac{1}{g_{m3} r_{o3}}\right] \tag{9.135}$$

Substituting for R_{im} and A_m from Eqs. (9.134) and (9.135), respectively, into Eq. (9.133), we obtain

$$G_{mcm} \simeq -\frac{1}{2R_{SS}}\,\frac{1}{g_{m3}(r_o \parallel r_{o3})} \tag{9.136}$$

where we have assumed that $\epsilon \ll 1$ and $g_{m3} r_{o3} \gg 1$.

Common-Mode Gain The common-mode gain can be found as

$$A_{cm} = G_{mcm}R_o$$
$$= -\frac{1}{2R_{SS}}\frac{r_{o2} \parallel r_{o4}}{g_{m3}(r_o \parallel r_{o3})}$$

Since $r_{o2} = r_o$ and $r_{o4} = r_{o3}$,

$$A_{cm} = -\frac{1}{2g_{m3}R_{SS}} \tag{9.137}$$

This expression clearly indicates that $|A_{cm}|$ is small and can be reduced by using a bias current source with a large output resistance R_{SS}.

The Common-Mode Rejection Ratio We can find the common-mode rejection ratio (CMRR) by dividing $|A_d|$ from Eq. (9.112) by $|A_{cm}|$ from Eq. (9.137). Alternatively, since both A_d and A_{cm} are obtained as the product of the respective G_m by R_o, we can find the CMRR by dividing $|G_{md}|$, which is equal to $g_{m1,2}$, by $|G_{mcm}|$ from Eq. (9.136):

$$\text{CMRR} = (2g_{m1,2}R_{SS})[g_{m3}(r_o \parallel r_{o3})] \tag{9.138}$$

In this expression $r_o = r_{o1,2} = R_{od}$ and $r_{o3} = r_{o4} = R_{om}$, thus

$$\text{CMRR} = [2g_{m1,2}R_{SS}]\,[g_{m3}(R_{od} \parallel R_{om})] \tag{9.139}$$

Note that the CMRR will be large. Its value can be increased by raising the value of the output resistance of the bias-current source, R_{SS}.

Video Example VE 9.4 Design of a Current-Mirror-Loaded MOS Differential Amplifier

Design the current-mirror-loaded differential MOS amplifier of Fig. 9.32(a) to obtain a differential gain of 50 V/V. The technology available provides $\mu_n C_{ox} = 4\mu_p C_{ox} = 400\,\mu\text{A/V}^2$, $|V_t| = 0.5$ V, and $|V_A'| = 20$ V/μm and operates from ±1-V supplies. Use a bias current $I = 200\ \mu$A and operate all devices at $|V_{OV}| =$ 0.2 V. Assume all transistors have the same channel length.

(a) Find the W/L ratios of the four transistors.
(b) Specify the channel length required of all transistors.
(c) If $V_{ICM} = 0$, what is the allowable range of v_O?
(d) If I is delivered by a simple NMOS current source operated at the same V_{OV} and having the same channel length as the other four transistors, determine the CMRR obtained.

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.



VE 9.4

Related end-of-chapter problem: 9.99

EXERCISE

9.19 For the current-mirror-loaded MOS differential amplifier specified in Exercise 9.17, let $R_{SS} = 25\ \text{k}\Omega$. Calculate $|A_{cm}|$ and CMRR. Use the results of Exercise 9.17.
Ans. 0.005 V/V; 20,000 or 86 dB

9.6 Multistage Amplifiers

Practical transistor amplifiers usually consist of a number of stages connected in cascade. In addition to providing gain, the first (or input) stage is usually required to provide a high input resistance in order to avoid loss of signal level when the amplifier is fed from a high-resistance source. In a differential amplifier the input stage must also provide large common-mode rejection. The function of the middle stages of an amplifier cascade is to provide the bulk of the voltage gain. In addition, the middle stages provide such other functions as the conversion of the signal from differential mode to single-ended mode (unless, of course, the amplifier output also is differential) and the shifting of the dc level of the signal in order to allow the output signal to swing both positive and negative. These two functions and others will be illustrated later in this section and in greater detail in Chapter 13.

Finally, the main function of the last (or output) stage of an amplifier is to provide a low output resistance in order to avoid loss of gain when a low-valued load resistance is connected to the amplifier. Also, the output stage should be able to supply the current required by the load in an efficient manner—that is, without dissipating an unduly large amount of power in the output transistors. We have already studied one type of amplifier configuration suitable for implementing output stages, namely, the source follower and the emitter follower. In Chapter 12 we will see that the source and emitter followers are not optimum from the point of view of power efficiency and that other, more appropriate circuit configurations exist for output stages that are required to supply large amounts of output power. In fact, we will encounter some such output stages in the op-amp circuit examples studied in Chapter 13.

To illustrate the circuit structure and the method of analysis of multistage amplifiers, we will present two examples: a two-stage CMOS op amp and a four-stage bipolar op amp.

THE INTERNATIONAL SOLID-STATE CIRCUITS CONFERENCE (ISSCC)

The International Solid-State Circuits Conference (ISSCC) is the foremost global forum for the presentation of novel developments in the broad field of solid-state circuits and systems. Sponsored by the Institute of Electrical and Electronics Engineers (IEEE), ISSCC began in 1954 at the University of Pennsylvania and later migrated to Silicon Valley where the ISSCC's annual meetings have been held in San Francisco since 1990.

In the early 1950s, shortly after the invention of the transistor, many of the large vacuum-tube-dependent electronics companies (e.g., GE, RCA) realized the potentially dramatic importance of these new amplifying devices. But eager as they were to create new transistor-based designs, they were quickly stymied by the lack of qualified engineers with appropriate background. In fact, Bell Labs was the only group with experience. The situation called for a novel solution, one in which limited expertise could be shared among multiple interested research and development entities. Thus was born the first "Transistor Conference," the precursor of ISSCC. Clearly, this was an idea that worked!

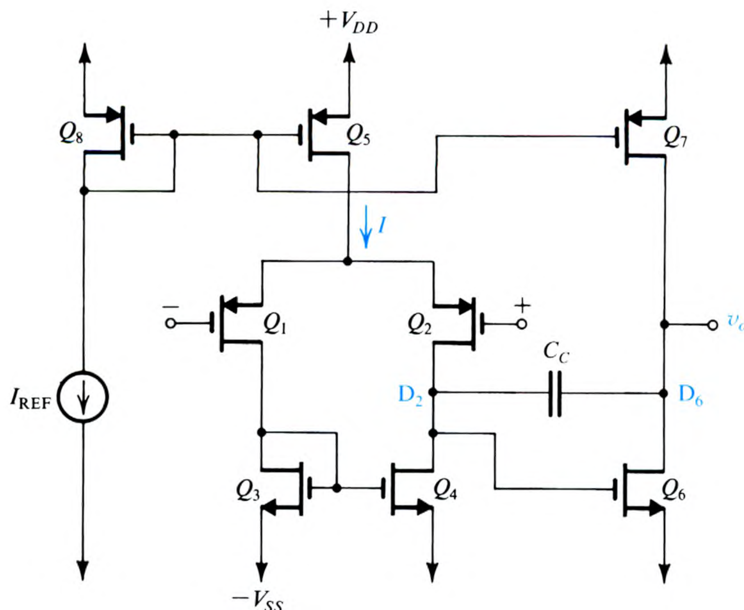


Figure 9.37 Two-stage CMOS op-amp configuration.

9.6.1 A Two-Stage CMOS Op Amp

Figure 9.37 shows a popular structure for CMOS op amps known as the **two-stage configuration**. The circuit utilizes two power supplies, which can range from ±0.9 V for the 0.18-μm technology down to ±0.5 V for the 65-nm technology. A reference bias current I_{REF} is generated either externally or using on-chip circuits. The current mirror formed by Q_8 and Q_5 supplies the differential pair Q_1–Q_2 with bias current. The W/L ratio of Q_5 is selected to yield the desired value for the input-stage bias current I (or $I/2$ for each of Q_1 and Q_2). The input differential pair is loaded with the current mirror formed by Q_3 and Q_4. Thus the input stage is identical to that studied in Section 9.5 (except that here the differential pair is implemented with PMOS transistors and the current mirror with NMOS).

The second stage consists of Q_6, which is a common-source amplifier loaded with the current-source transistor Q_7. A capacitor C_C is included in the negative-feedback path of the second stage. Its function will be explained in Chapter 13, when we study the frequency response of this amplifier.

A striking feature of the circuit in Fig. 9.37 is that it does *not* have a low-output-resistance stage. In fact, the output resistance of the circuit is equal to $(r_{o6} \| r_{o7})$ and is thus rather high. This circuit, therefore, is not suitable for driving low-impedance loads. Nevertheless, the circuit is very popular and is used frequently for implementing op amps in VLSI circuits, where the op amp needs to drive only a small capacitive load, for example, in switched-capacitor circuits (Chapter 14). Also, when this op amp is utilized, negative feedback is applied, which results in reducing the output resistance, as will be seen in Chapter 11. The simplicity of the circuit results in an op amp of reasonably good quality realized in a very small chip area.

Voltage Gain The differential voltage gain of the first stage was found in Section 9.5 to be given by

$$A_1 = -g_{m1}(r_{o2} \| r_{o4}) \tag{9.140}$$

where g_{m1} is the transconductance of each of the transistors of the first stage, that is, Q_1 and Q_2.

The second stage is a current-source-loaded, common-source amplifier whose voltage gain is given by

$$A_2 = -g_{m6}(r_{o6} \| r_{o7}) \tag{9.141}$$

The dc open-loop gain of the op amp is the product of A_1 and A_2.

Example 9.6

Consider the circuit in Fig. 9.37 with the following device geometries (in μm).

Transistor	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8
W/L	10/0.4	10/0.4	2.5/0.4	2.5/0.4	20/0.4	5/0.4	20/0.4	20/0.4

Let $I_{REF} = 100\ \mu A$, $V_{tn} = 0.5$ V, $V_{tp} = -0.5$ V, $\mu_n C_{ox} = 400\ \mu A/V^2$, $\mu_p C_{ox} = 100\ \mu A/V^2$, $|V_A|$ (for all devices) = 5 V, $V_{DD} = V_{SS} = 1$ V. For all devices, evaluate I_D, $|V_{OV}|$, $|V_{GS}|$, g_m, and r_o. Also find A_1, A_2, the dc open-loop voltage gain, the input common-mode range, and the output voltage range. Neglect the effect of V_A on bias current.

Solution

Refer to Fig. 9.37. Since Q_8 and Q_5 are matched, $I = I_{REF}$. Thus Q_1, Q_2, Q_3, and Q_4 each conducts a current equal to $I/2 = 50\ \mu A$. Since Q_7 is matched to Q_5 and Q_8, the current in Q_7 is equal to $I_{REF} = 100\ \mu A$. Finally, Q_6 conducts an equal current of 100 μA.

With I_D of each device known, we use

$$I_D = \frac{1}{2}(\mu C_{ox})(W/L)V_{OV}^2$$

to determine $|V_{OV}|$ for each transistor. Then we find $|V_{GS}|$ from $|V_{GS}| = |V_t| + |V_{OV}|$. The results are given in Table 9.1.

The transconductance of each device is determined from

$$g_m = 2I_D/|V_{OV}|$$

The value of r_o is determined from

$$r_o = |V_A|/I_D$$

The resulting values of g_m and r_o are given in Table 9.1.

Table 9.1

	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8
I_D (μA)	50	50	50	50	100	100	100	100
$\|V_{OV}\|$(V)	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
$\|V_{GS}\|$(V)	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
g_m (mA/V)	0.5	0.5	0.5	0.5	1	1	1	1
r_o (kΩ)	100	100	100	100	50	50	50	50

The voltage gain of the first stage is determined from

$$A_1 = -g_{m1}(r_{o2} \| r_{o4})$$
$$= -0.5(100 \| 100) = -25 \text{ V/V}$$

The voltage gain of the second stage is determined from

$$A_2 = -g_{m6}(r_{o6} \| r_{o7})$$
$$= -1(50 \| 50) = -25 \text{V/V}$$

Thus the overall dc open-loop gain is

$$A_0 = A_1 A_2 = (-25) \times (-25) = 625 \text{ V/V}$$

or

$$20 \log 625 = 56 \text{ dB}$$

The lower limit of the input common-mode range is the value of input voltage at which Q_1 and Q_2 leave the saturation region. This occurs when the input voltage falls below the voltage at the drain of Q_1 by $|V_{tp}|$ volts. Since the drain of Q_1 is at $-1.0 + 0.7 = -0.3$ V, then the lower limit of the input common-mode range is -0.8 V.

The upper limit of the input common-mode range is the value of input voltage at which Q_5 leaves the saturation region. Since for Q_5 to operate in saturation the voltage across it (i.e., V_{SD5}) should at least be equal to the overdrive voltage at which it is operating (i.e., 0.2 V), the highest voltage permitted at the drain of Q_5 should be $+0.8$ V. It follows that the highest value of V_{ICM} should be

$$V_{ICM\max} = 0.8 - 0.7 = +0.1 \text{V}$$

The highest allowable output voltage is the value at which Q_7 leaves the saturation region, which is $V_{DD} - |V_{OV7}| = 1.0 - 0.2 = 0.8$ V. The lowest allowable output voltage is the value at which Q_6 leaves saturation, which is $-V_{SS} + V_{OV6} = -1.0 + 0.2 = -0.8$ V. Thus, the output voltage range is -0.8 V to $+0.8$ V.

Input Offset Voltage The device mismatches inevitably present in the input stage give rise to an input offset voltage. The components of this input offset voltage can be calculated

using the methods developed in Section 9.4.1. Because device mismatches are random, the resulting offset voltage is referred to as **random offset**. This is to distinguish it from another type of input offset voltage that can be present even if all appropriate devices are perfectly matched. This predictable or **systematic offset** can be minimized by careful design. Although it occurs also in BJT op amps, it is usually much more pronounced in CMOS op amps because their gain-per-stage is rather low.

To see how systematic offset can occur in the circuit of Fig. 9.37, let the two input terminals be grounded. If the input stage is perfectly balanced, then the voltage appearing at the drain of Q_4 will be equal to that at the drain of Q_3, which is $(-V_{SS}+V_{GS4})$. Now this is also the voltage that is fed to the gate of Q_6. In other words, a voltage equal to V_{GS4} appears between gate and source of Q_6. Thus the drain current of Q_6, I_6, will be related to the drain current of Q_4, which is equal to $I/2$, by the relationship

$$I_6 = \frac{(W/L)_6}{(W/L)_4}(I/2) \tag{9.142}$$

In order for no offset voltage to appear at the output, this current must be exactly equal to the current supplied by Q_7. The latter current is related to the current I of the parallel transistor Q_5 by

$$I_7 = \frac{(W/L)_7}{(W/L)_5}I \tag{9.143}$$

Now, the condition for making $I_6 = I_7$ can be found from Eqs. (9.142) and (9.143) as

$$\frac{(W/L)_6}{(W/L)_4} = 2\frac{(W/L)_7}{(W/L)_5} \tag{9.144}$$

If this condition is not met, a systematic offset will result. From the specification of the device geometries in Example 9.6, we can verify that condition (9.144) is satisfied, and, therefore, the op amp analyzed in that example should not exhibit a systematic input offset voltage.

EXERCISE

9.20 Consider the CMOS op amp of Fig. 9.37 when fabricated in a 0.13-μm CMOS technology for which $\mu_n C_{ox} = 4\mu_p C_{ox} = 512\ \mu\text{A/V}^2$, $|V_t| = 0.4$ V, and $V_{DD} = V_{SS} = 0.65$ V. For a particular design, $I = 128\ \mu\text{A}$, $(W/L)_1 = (W/L)_2 = \frac{1}{2}(W/L)_5 = 44.4$, and $(W/L)_3 = (W/L)_4 = 11.1$.

(a) Find the (W/L) ratios of Q_6 and Q_7 so that $I_6 = 128\ \mu\text{A}$.
(b) Find the overdrive voltage, $|V_{OV}|$, at which each of Q_1, Q_2, and Q_6 is operating.
(c) Find g_m for Q_1, Q_2, and Q_6.
(d) If $|V_A| = 3.2$ V, find r_{o2}, r_{o4}, r_{o6}, and r_{o7}.
(e) Find the voltage gains A_1 and A_2, and the overall gain A.

Ans. (a) $(W/L)_6 = 22.2, (W/L)_7 = 88.8$; (b) 0.15 V, 0.15 V, 0.15 V; (c) 0.853 mA/V, 0.853 mA/V, 1.71 mA/V; (d) 50 kΩ, 50 kΩ, 25 kΩ, 25 kΩ; (e) –21.3 V/V, –21.3 V/V, 455 V/V

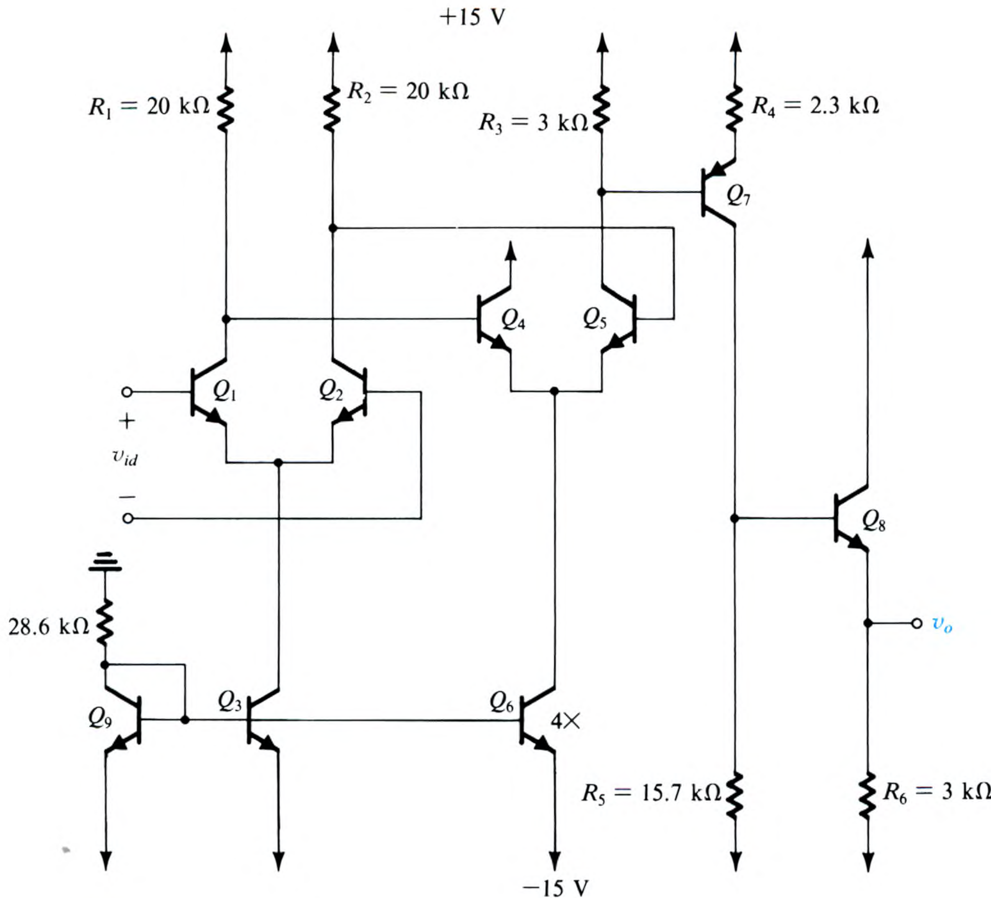


Figure 9.38 A four-stage bipolar op amp.

9.6.2 A Bipolar Op Amp

Our second example of multistage amplifiers is the bipolar op amp shown in Fig. 9.38. The circuit consists of four stages. The **differential-in, differential-out** input stage consists of transistors Q_1 and Q_2, which are biased by current source Q_3. The second stage is also a differential-input amplifier, but its output is taken single-endedly at the collector of Q_5. This stage is formed by Q_4 and Q_5, which are biased by the current source Q_6. Note that the conversion from differential to single-ended as performed by the second stage results in a loss of gain by a factor of 2. In the more elaborate method for accomplishing this conversion studied in Section 9.5, a current mirror was used as an active load.

In addition to providing some voltage gain, the third stage, consisting of the *pnp* transistor Q_7, provides the essential function of *shifting the dc level* of the signal. Thus, while the signal at the collector of Q_5 is not allowed to swing below the voltage at the base of Q_5 (+10 V), the signal at the collector of Q_7 can swing negatively (and positively, of course). From our study of op amps in Chapter 2, we know that the output terminal of the op amp should be capable of both positive and negative voltage swings. Therefore every op-amp circuit includes a **level-shifting** arrangement. Although the use of the complementary *pnp* transistor provides a simple solution to the level-shifting problem, other forms of level shifter

exist. As an example, note that level shifting is accomplished in the CMOS op amp we have studied in Section 9.6.1 by using complementary devices for the two stages: that is, p-channel for the first stage and n-channel for the second stage.

The output stage of the op amp consists of emitter follower Q_8. As we know from our study of op amps in Chapter 2, ideally the output operates around zero volts. This and other features of the BJT op amp will be illustrated in Example 9.7.

Example 9.7

In this example, we analyze the dc bias of the bipolar op-amp circuit of Fig. 9.38. To that end, Fig. 9.39 shows the circuit with the two input terminals connected to ground.

(a) Perform an approximate dc analysis (assuming $\beta \gg 1$, $|V_{BE}| \simeq 0.7$ V, and neglecting the Early effect) to calculate the dc currents and voltages everywhere in the circuit. Note that Q_6 has four times the area of each of Q_9 and Q_3.

(b) Calculate the quiescent power dissipation in this circuit.

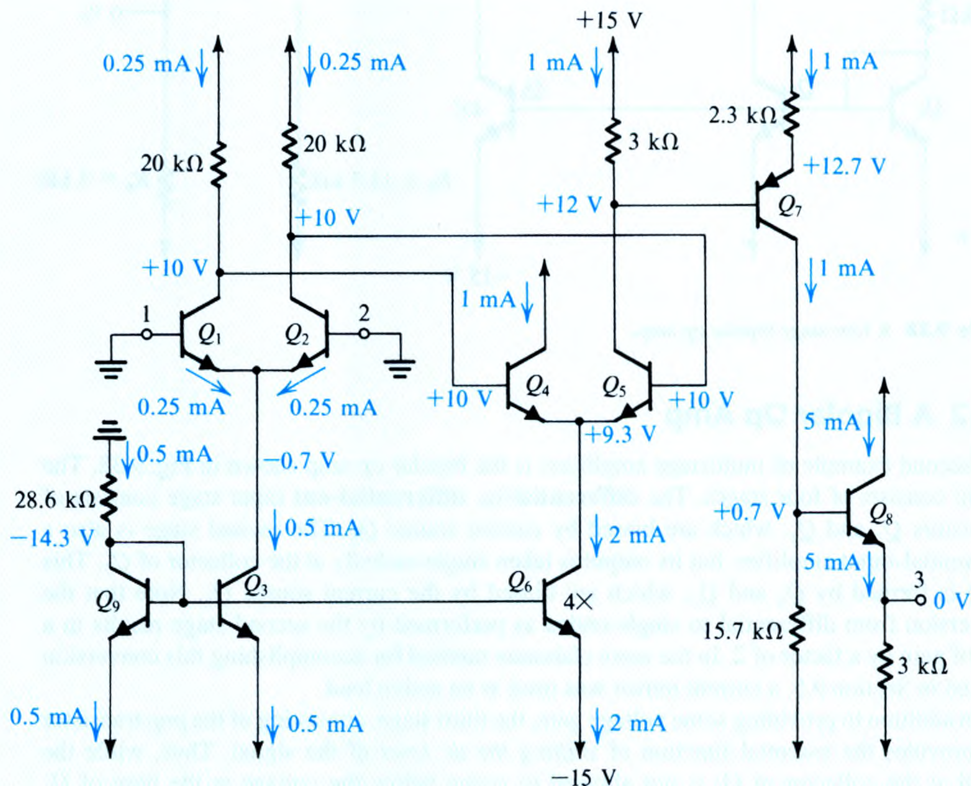


Figure 9.39 Circuit for Example 9.7 with all the dc currents and voltages shown.

(c) If transistors Q_1 and Q_2 have $\beta = 100$, calculate the input bias current of the op amp.
(d) What is the input common-mode range of this op amp?

Solution

(a) The values of all dc currents and voltages are shown on the circuit diagram. These values were calculated by ignoring the base current of every transistor—that is, by assuming β to be very high. We start by determining the current through the diode-connected transistor Q_9 to be 0.5 mA. Then we see that transistor Q_3 conducts 0.5 mA and transistor Q_6 conducts 2 mA. The current-source transistor Q_3 feeds the differential pair (Q_1, Q_2) with 0.5 mA. Thus each of Q_1 and Q_2 will be biased at 0.25 mA. The collectors of Q_1 and Q_2 will be at $[+15 - 0.25 \times 20] = +10$ V.

In the second differential stage formed by Q_4 and Q_5, we find the voltage at their emitters to be $[+10 - 0.7] = 9.3$ V. This differential pair is biased by the current-source transistor Q_6, which supplies a current of 2 mA; thus Q_4 and Q_5 will each be biased at 1 mA. We can now calculate the voltage at the collector of Q_5 as $[+15 - 1 \times 3] = +12$ V. This will cause the voltage at the emitter of the *pnp* transistor Q_7 to be +12.7 V, and the emitter current of Q_7 will be $(+15 - 12.7)/2.3 = 1$ mA.

The collector current of Q_7, 1 mA, causes the voltage at the collector to be $[-15 + 1 \times 15.7] = +0.7$ V. The emitter of Q_8 will be 0.7 V below the base; thus output terminal 3 will be at 0 V. Finally, the emitter current of Q_8 can be calculated to be $[0 - (-15)]/3 = 5$ mA.

(b) To calculate the power dissipated in the circuit in the quiescent state (i.e., with zero input signal) we evaluate the dc current the circuit draws from each of the two power supplies. From the +15 V supply the dc current is $I^+ = 0.25 + 0.25 + 1 + 1 + 1 + 5 = 8.5$ mA; thus the power supplied by the positive supply is $P^+ = 15 \times 8.5 = 127.5$ mW. The –15-V supply provides a current I^- given by $I^- = 0.5 + 0.5 + 2 + 1 + 5 = 9$ mA; thus the power provided by the negative supply is $P^- = 15 \times 9 = 135$ mW. Adding P^+ and P^- gives us the total power dissipated in the circuit P_D: $P_D = P^+ + P^- = 262.5$ mW.

(c) The input bias current of the op amp is the average of the dc currents that flow in the two input terminals (i.e., in the bases of Q_1 and Q_2). These two currents are equal (because we have assumed matched devices); thus the bias current is given by

$$I_B = \frac{I_{E1}}{\beta + 1} \simeq 2.5\ \mu\text{A}$$

(d) The upper limit on the input common-mode voltage is determined by the voltage at which Q_1 and Q_2 leave the active mode and enter saturation. This will happen if the input voltage exceeds the collector voltage, which is +10 V, by about 0.4 V. Thus the upper limit of the common-mode range is +10.4 V.

The lower limit of the input common-mode range is determined by the voltage at which Q_3 leaves the active mode and thus ceases to act as a constant-current source. This will happen if the collector voltage of Q_3 goes below the voltage at its base, which is –14.3 V, by more than 0.4 V. It follows that the input common-mode voltage should not go lower than $-14.7 + 0.7 = -14$ V. Thus the common-mode range is –14 V to +10.4 V.

Example 9.8

Use the dc bias quantities evaluated in Example 9.7 to analyze the circuit in Fig. 9.38 to determine the input resistance, the voltage gain, and the output resistance.

Solution

The input differential resistance R_{id} is given by

$$R_{id} = r_{\pi 1} + r_{\pi 2}$$

Since each of Q_1 and Q_2 is operating at an emitter current of 0.25 mA, it follows that

$$r_{e1} = r_{e2} = \frac{25}{0.25} = 100\ \Omega$$

Assume $\beta = 100$; then

$$r_{\pi 1} = r_{\pi 2} = 101 \times 100 = 10.1\ \text{k}\Omega$$

Thus,

$$R_{id} = 20.2\ \text{k}\Omega$$

To evaluate the gain of the first stage, we first find the input resistance of the second stage, R_{i2},

$$R_{i2} = r_{\pi 4} + r_{\pi 5}$$

Q_4 and Q_5 are each operating at an emitter current of 1 mA; thus

$$r_{e4} = r_{e5} = 25\ \Omega$$
$$r_{\pi 4} = r_{\pi 5} = 101 \times 25 = 2.525\ \text{k}\Omega$$

Thus $R_{i2} = 5.05\ \text{k}\Omega$. This resistance appears between the collectors of Q_1 and Q_2, as shown in Fig. 9.40. Thus the gain of the first stage will be

$$A_1 \equiv \frac{v_{o1}}{v_{id}} \simeq \frac{\text{Total resistance in collector circuit}}{\text{Total resistance in emitter circuit}}$$

$$= \frac{R_{i2} \parallel (R_1 + R_2)}{r_{e1} + r_{e2}}$$

$$= \frac{5.05\ \text{k}\Omega \parallel 40\ \text{k}\Omega}{200\ \Omega} = 22.4\ \text{V/V}$$

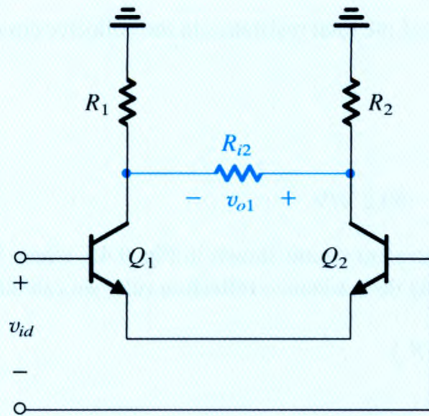


Figure 9.40 Equivalent circuit for calculating the gain of the input stage of the amplifier in Fig. 9.38.

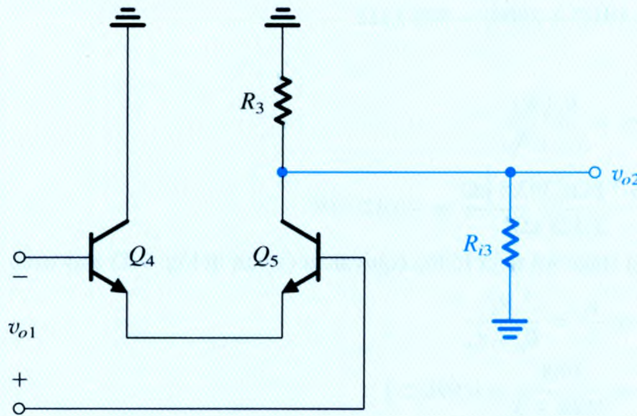


Figure 9.41 Equivalent circuit for calculating the gain of the second stage of the amplifier in Fig. 9.38.

Figure 9.41 shows an equivalent circuit for calculating the gain of the second stage. As indicated, the input voltage to the second stage is the output voltage of the first stage, v_{o1}. Also shown is the resistance R_{i3}, which is the input resistance of the third stage formed by Q_7. The value of R_{i3} can be found by multiplying the total resistance in the emitter of Q_7 by $(\beta + 1)$:

$$R_{i3} = (\beta + 1)(R_4 + r_{e7})$$

Since Q_7 is operating at an emitter current of 1 mA,

$$r_{e7} = \frac{25}{1} = 25\ \Omega$$

$$R_{i3} = 101 \times 2.325 = 234.8\ \text{k}\Omega$$

Example 9.8 *continued*

We can now find the gain A_2 of the second stage as the ratio of the total resistance in the collector circuit to the total resistance in the emitter circuit:

$$A_2 \equiv \frac{v_{o2}}{v_{o1}} \simeq -\frac{R_3 \| R_{i3}}{r_{e4} + r_{e5}}$$

$$= -\frac{3\text{ k}\Omega \| 234.8\text{ k}\Omega}{50\ \Omega} = -59.2\text{ V/V}$$

To obtain the gain of the third stage we refer to the equivalent circuit shown in Fig. 9.42, where R_{i4} is the input resistance of the output stage formed by Q_8. Using the resistance-reflection rule, we calculate the value of R_{i4} as

$$R_{i4} = (\beta + 1)(r_{e8} + R_6)$$

where

$$r_{e8} = \frac{25}{5} = 5\ \Omega$$

$$R_{i4} = 101(5 + 3000) = 303.5\text{ k}\Omega$$

The gain of the third stage is given by

$$A_3 \equiv \frac{v_{o3}}{v_{o2}} \simeq -\frac{R_5 \| R_{i4}}{r_{e7} + R_4}$$

$$= -\frac{15.7\text{ k}\Omega \| 303.5\text{ k}\Omega}{2.325\text{ k}\Omega} = -6.42\text{ V/V}$$

Finally, to obtain the gain A_4 of the output stage we refer to the equivalent circuit in Fig. 9.43 and write

$$A_4 \equiv \frac{v_o}{v_{o3}} = \frac{R_6}{R_6 + r_{e8}}$$

$$= \frac{3000}{3000 + 5} = 0.998 \simeq 1$$

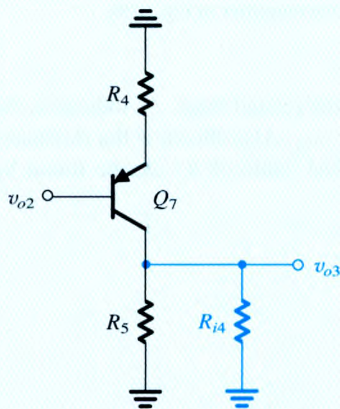


Figure 9.42 Equivalent circuit for evaluating the gain of the third stage in the amplifier circuit of Fig. 9.38.

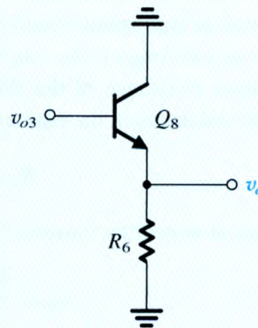


Figure 9.43 Equivalent circuit of the output stage of the amplifier circuit of Fig. 9.38.

The overall voltage gain of the amplifier can then be obtained as follows:

$$\frac{v_o}{v_{id}} = A_1A_2A_3A_4 = 8513 \text{ V/V}$$

or 78.6 dB.

To obtain the output resistance R_o we "grab hold" of the output terminal in Fig. 9.38 and look back into the circuit. By inspection we find

$$R_o = R_6 \| \left[r_{e8} + R_5/(\beta + 1)\right]$$

which gives

$$R_o = 152\ \Omega$$

EXERCISE

9.21 Use the results of Example 9.8 to calculate the overall voltage gain of the amplifier in Fig. 9.38 when it is connected to a source having a resistance of 10 kΩ and a load of 1 kΩ.
Ans. 4943 V/V

Analysis Using Current Gains There is an alternative method for analyzing bipolar multistage amplifiers that can be easier to perform in some cases. The method makes use of current gains or more appropriately current-transmission factors. In effect, we trace the transmission of the signal current throughout the amplifier cascade, evaluating all the current transmission factors in turn. We shall illustrate the method by using it to analyze the amplifier circuit of the preceding example.

Figure 9.44 shows the amplifier circuit prepared for small-signal analysis. We have indicated the signal currents through all the circuit branches. Also shown are the input resistances of all four stages of the amplifier. These should be evaluated before we begin the analysis.

The purpose of the analysis is to determine the overall voltage gain (v_o/v_{id}). To that end, we express v_o in terms of the signal current in the emitter of Q_8, i_{e8}, and v_{id} in terms of the input signal current i_i, as follows:

$$v_o = R_6 i_{e8}$$
$$v_{id} = R_{i1} i_i$$

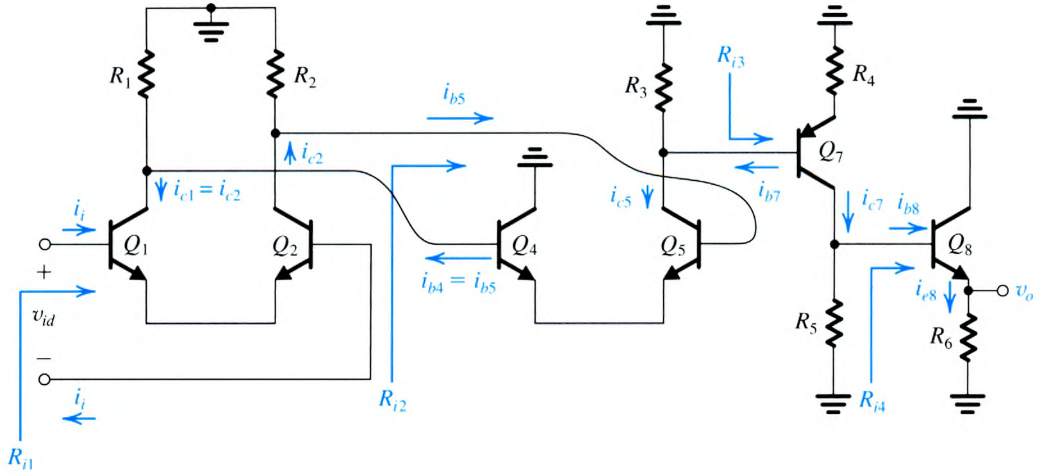


Figure 9.44 The circuit of the multistage amplifier of Fig. 9.38 prepared for small-signal analysis. Indicated are the signal currents throughout the amplifier and the input resistances of the four stages.

Thus, the voltage gain can be expressed in terms of the current gain (i_{e8}/i_i) as

$$\frac{v_o}{v_{id}} = \frac{R_6}{R_{i1}}\frac{i_{e8}}{i_i}$$

Next, we expand the current gain (i_{e8}/i_i) in terms of the signal currents throughout the circuit as follows:

$$\frac{i_{e8}}{i_i} = \frac{i_{e8}}{i_{b8}} \times \frac{i_{b8}}{i_{c7}} \times \frac{i_{c7}}{i_{b7}} \times \frac{i_{b7}}{i_{c5}} \times \frac{i_{c5}}{i_{b5}} \times \frac{i_{b5}}{i_{c2}} \times \frac{i_{c2}}{i_i}$$

Each of the current-transmission factors on the right-hand side is either the current gain of a transistor or the ratio of a current divider. Thus, reference to Fig. 9.44 enables us to find these factors by inspection:

$$\frac{i_{e8}}{i_{b8}} = \beta_8 + 1 \qquad \frac{i_{b8}}{i_{c7}} = \frac{R_5}{R_5 + R_{i4}}$$

$$\frac{i_{c7}}{i_{b7}} = \beta_7 \qquad \frac{i_{b7}}{i_{c5}} = \frac{R_3}{R_3 + R_{i3}}$$

$$\frac{i_{c5}}{i_{b5}} = \beta_5 \qquad \frac{i_{b5}}{i_{c2}} = \frac{(R_1 + R_2)}{(R_1 + R_2) + R_{i2}}$$

$$\frac{i_{c2}}{i_i} = \beta_2$$

These ratios can be easily evaluated and their values used to determine the voltage gain.

With a little practice, it is possible to carry out such an analysis very quickly, without explicitly labeling the signal currents on the circuit diagram. We can simply "walk through"

the circuit, from input to output, or vice versa, determining the current-transmission factors one at a time, in a chainlike fashion.

EXERCISE

9.22 Use the values of input resistance found in Example 9.8 to evaluate the seven current-transmission factors and hence the overall current gain and voltage gain.
Ans. The current-transmission factors in the order of their listing are 101, 0.0492, 100, 0.0126, 100, 0.8879, 100 A/A; the overall current gain is 55599 A/A; the voltage gain is 8257 V/V. This value differs slightly from that found in Example 9.8, because of the various approximations made in the example (e.g., $\alpha \simeq 1$).

Summary

- The differential-pair or differential-amplifier configuration is the most widely used building block in analog IC design. The input stage of every op amp is a differential amplifier.
- There are two reasons for preferring differential to single-ended amplifiers: Differential amplifiers are insensitive to interference, and they do not need bypass and coupling capacitors.
- For a MOS (bipolar) pair biased by a current source I, each device operates at a drain (collector, assuming $\alpha = 1$) current of $I/2$ and a corresponding overdrive voltage V_{OV} (no counterpart in bipolar). Each device has $g_m = I/V_{OV}$ ($\alpha I/2V_T$, for bipolar) and $r_o = |V_A|/(I/2)$.
- With the two input terminals connected to a suitable dc voltage V_{CM}, the bias current I of a perfectly symmetrical differential pair divides equally between the two transistors, resulting in a zero voltage difference between the two drains (collectors). To steer the current completely to one side of the pair requires a difference input voltage v_{id} of at least $\sqrt{2}V_{OV}$ ($4V_T$ for bipolar).
- Superimposing a differential input signal v_{id} on the dc common-mode input voltage V_{CM} such that $v_{I1} = V_{CM} + v_{id}/2$ and $v_{I2} = V_{CM} - v_{id}/2$ causes a virtual signal ground to appear on the common-source (common-emitter) connection. In response to v_{id}, the current in Q_1 increases by $g_m v_{id}/2$ and the current in Q_2 decreases by $g_m v_{id}/2$. Thus, voltage signals of $\pm g_m(R_D \parallel r_o)v_{id}/2$ develop at the two drains (collectors, with R_D replaced by R_C). If the output voltage is taken single-endedly, that is, between one of the drains (collectors) and ground, a differential gain of $\frac{1}{2}g_m(R_D \parallel r_o)$ is realized. When the output is taken differentially, that is, between the two drains (collectors), the differential gain realized is twice as large: $g_m(R_D \parallel r_o)$.
- The analysis of a differential amplifier to determine differential gain, differential input resistance, frequency response of differential gain, and so on is simplified by using the differential half-circuit, a common-source (common-emitter) transistor biased at $I/2$.
- An input common-mode signal v_{icm} gives rise to drain (collector) voltage signals that are ideally equal and given by $-v_{icm}(R_D/2R_{SS})[-v_{icm}(R_C/2R_{EE})$ for the bipolar pair], where R_{SS} (R_{EE}) is the output resistance of the current source that supplies the bias current I. When the output is taken single-endedly, a common-mode gain of magnitude $|A_{cm}| = R_D/2R_{SS}$ ($R_C/2R_{EE}$ for the bipolar case) results. Taking the output differentially results, in the perfectly matched case, in zero A_{cm} (infinite CMRR). Mismatches between the two sides of the pair make A_{cm} finite even when the output is taken differentially: A mismatch ΔR_D causes $|A_{cm}| = (R_D/2R_{SS})(\Delta R_D/R_D)$; a mismatch Δk_n causes $|A_{cm}| = (R_D/2R_{SS})$ $(\Delta k_n/k_n)$. Corresponding expressions apply for the bipolar pair.
- While the input differential resistance R_{id} of the MOS pair is infinite, that for the bipolar pair is only $2r_\pi$ but can

be increased to $2(\beta+1)(r_e+R_e)$ by including resistances R_e in the two emitters. This will lower A_d, however.

- Mismatches between the two sides of a differential pair result in a differential dc output voltage V_O even when the two input terminals are tied together and connected to a dc voltage V_{CM}. This signifies the presence of an input offset voltage. In a MOS pair there are three main sources for V_{OS}:

$$\Delta R_D \Rightarrow V_{OS} = \frac{V_{OV}}{2}\frac{\Delta R_D}{R_D}$$

$$\Delta k_n \Rightarrow V_{OS} = \frac{V_{OV}}{2}\frac{\Delta k_n}{k_n}$$

$$\Delta V_t \Rightarrow V_{OS} = \Delta V_t$$

For the bipolar pair there are two main sources:

$$\Delta R_C \Rightarrow V_{OS} = V_T\frac{\Delta R_C}{R_C}$$

$$\Delta I_S \Rightarrow V_{OS} = V_T\frac{\Delta I_S}{I_S}$$

- A popular circuit in both MOS and bipolar analog ICs is the current-mirror-loaded differential pair. It realizes a high differential gain $A_d = g_m(R_{o\ \text{pair}} \| R_{o\ \text{mirror}})$ and a low common-mode gain, $|A_{cm}| = 1/2g_{m3}R_{SS}$ for the MOS circuit, as well as performing the differential-to-single-ended conversion with no loss of gain.
- The CMOS two-stage amplifier studied in Section 9.6.1 is intended for use as part of an IC system and thus is required to drive only small capacitive loads. Therefore it does not have an output stage with a low output resistance.
- A multistage amplifier typically consists of three or more stages: an input stage having a high input resistance, a reasonably high gain, and, if differential, a high CMRR; one or two intermediate stages that realize the bulk of the gain; and an output stage having a low output resistance. In designing and analyzing a multistage amplifier, the loading effect of each stage on the one that precedes it must be taken into account.

PROBLEMS

Computer Simulation Problems

SIM Problems identified by this icon are intended to demonstrate the value of using SPICE simulation to verify hand analysis and design, and to investigate important issues such as allowable signal swing and amplifier nonlinear distortion. Instructions to assist in setting up simulations for all the indicated problems can be found in the corresponding files on the website. Note that if a particular parameter value is not specified in the problem statement, you are to make a reasonable assumption.

Section 9.1: The MOS Differential Pair

9.1 For an NMOS differential pair with a common-mode voltage V_{CM} applied, as shown in Fig. 9.2, let $V_{DD} = V_{SS} = 1.0$ V, $k_n' = 0.4$ mA/V^2, $(W/L)_{1,2} = 10$, $V_{tn} = 0.4$ V, $I = 0.16$ mA, $R_D = 5$ kΩ, and neglect channel-length modulation.

(a) Find V_{OV} and V_{GS} for each transistor.
(b) For $V_{CM} = 0$, find V_S, I_{D1}, I_{D2}, V_{D1}, V_{D2}, and V_O.
(c) Repeat (b) for $V_{CM} = +0.4$ V.
(d) Repeat (b) for $V_{CM} = -0.1$ V.
(e) What is the highest value of V_{CM} for which Q_1 and Q_2 remain in saturation?
(f) If current source I requires a minimum voltage of 0.2 V to operate properly, what is the lowest value allowed for V_S and hence for V_{CM}?
(g) What is the input common-mode range?

9.2 For the PMOS differential amplifier shown in Fig. P9.2, let $V_{tp} = -0.4$ V and $k_p' W/L = 5$ mA/V^2. Neglect channel-length modulation.

problems with blue numbers are considered essential; * = difficult problem; ** = more difficult; *** = very challenging
SIM = simulation; D = design problem; ● = see related video example

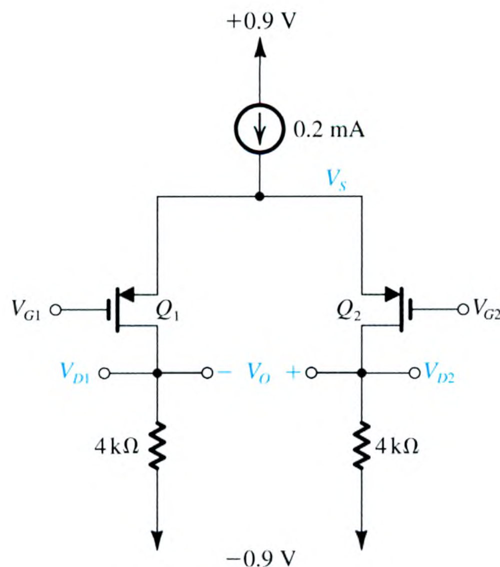


Figure P9.2

(a) For $V_{G1} = V_{G2} = 0$ V, find $|V_{OV}|$ and V_{SG} for each of Q_1 and Q_2. Also find V_S, V_{D1}, V_{D2}, and V_O.

(b) If the current source requires a minimum voltage of 0.2 V, find the input common-mode range.

9.3 For the differential amplifier specified in Problem 9.1 let $v_{G2} = 0$ and $v_{G1} = v_{id}$. Find the value of v_{id} that corresponds to each of the following situations:
(a) $i_{D1} = i_{D2} = 0.08$ mA; (b) $i_{D1} = 0.12$ mA and $i_{D2} = 0.04$ mA; (c) $i_{D1} = 0.16$ mA and $i_{D2} = 0$ (Q_2 just cuts off); (d) $i_{D1} = 0.04$ mA and $i_{D2} = 0.12$ mA; (e) $i_{D1} = 0$ mA (Q_1 just cuts off) and $i_{D2} = 0.16$ mA. For each case, find v_S, v_{D1}, v_{D2}, and v_O.

SIM **9.4** For the differential amplifier specified in Problem 9.2, let $v_{G2} = 0$ and $v_{G1} = v_{id}$. Find the range of v_{id} needed to steer the bias current from one side of the pair to the other. At each end of this range, give the value of the voltage at the common-source terminal, the drain voltages, and v_O.

9.5 Consider the differential amplifier specified in Problem 9.1 with G_2 grounded and $v_{G1} = v_{id}$. Let v_{id} be adjusted to the value that causes $i_{D1} = 0.09$ mA and $i_{D2} = 0.07$ mA. Find the corresponding values of v_{GS2}, v_S, v_{GS1}, and hence v_{id}. What is the difference output voltage v_O? What is the voltage gain v_O/v_{id}? What value of v_{id} results in $i_{D1} = 0.07$ mA and $i_{D2} = 0.09$ mA?

D 9.6 Design the circuit in Fig. P9.6 to obtain a dc voltage of 0 V at each of the drains of Q_1 and Q_2 when $v_{G1} = v_{G2} = 0$ V. Operate all transistors at $V_{OV} = 0.15$ V

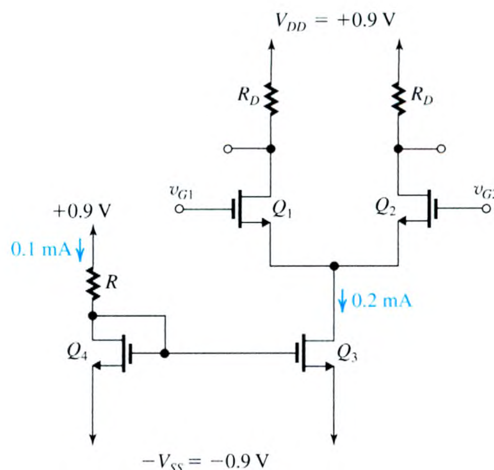


Figure P9.6

and assume that for the process technology in which the circuit is fabricated, $V_{tn} = 0.35$ V and $\mu_n C_{ox} = 400\ \mu\text{A/V}^2$. Neglect channel-length modulation. Determine the values of R, R_D, and the W/L ratios of Q_1, Q_2, Q_3, and Q_4. What is the input common-mode voltage range for your design?

9.7 The table providing the answers to Exercise 9.3 shows that as the maximum input signal to be applied to the differential pair is increased, linearity is maintained at the same level by operating at a higher V_{OV}. If $|v_{id}|_{max}$ is to be 230 mV, use the data in the table to determine the required V_{OV} and the corresponding values of W/L and g_m.

9.8 Use Eq. (9.23) to show that if the term involving v_{id}^2 is to be kept to a maximum value of k then the maximum possible fractional change in the transistor current is given by

$$\frac{\Delta I_{max}}{I/2} = 2\sqrt{k(1-k)}$$

and the corresponding maximum value of v_{id} is given by

$$v_{id\max} = 2\sqrt{k}V_{OV}$$

Evaluate both expressions for $k = 0.01$, 0.1, and 0.2.

9.9 A MOS differential amplifier biased with a current source $I = 100\ \mu$A is found to switch currents completely to one side of the pair when a difference signal $v_{id} = 0.25$ V is applied. At what overdrive voltage will each of Q_1 and Q_2 be operating when $v_{id} = 0$? If v_{id} for full current switching is to be 0.5 V, what must the bias current I be changed to?

D 9.10 Design the MOS differential amplifier of Fig. 9.5 to operate at $V_{OV} = 0.2$ V and to provide a transconductance g_m of 2 mA/V. Specify the W/L ratios and the bias

current. The technology available provides $V_t = 0.5$ V and $\mu_n C_{ox} = 400\ \mu\text{A/V}^2$.

***9.11** For the MOS differential pair in Fig. 9.5, specify the value of $v_{id} \equiv v_{G1} - v_{G2}$, in terms of V_{OV}, that

(a) causes i_{D1} to increase by 10% above its equilibrium value of $I/2$.
(b) makes $i_{D1}/i_{D2} = 1.0; 2.0; 1.1; 1.01; 20$.

9.12 An NMOS differential amplifier is operated at a bias current I of 0.2 mA and has a W/L ratio of 12.5, $\mu_n C_{ox} = 400\ \mu\text{A/V}^2$, $V_A = 5$ V, and $R_D = 20$ kΩ. Find V_{OV}, g_m, r_o, and A_d.

D 9.13 Design an NMOS differential amplifier to operate with a differential input voltage that can be as high as 0.1 V while keeping the nonlinear term under the square root in Eq. (9.23) to a maximum of 0.04. A transconductance g_m of 2 mA/V is needed, and the amplifier must provide a differential output signal of 1 V when the input is at its maximum value. Find the required values of V_{OV}, I, R_D, and W/L. Assume that the technology available has $\mu_n C_{ox} = 200\ \mu\text{A/V}^2$ and $\lambda = 0$.

D 9.14 Design a MOS differential amplifier to operate from ±1-V power supplies and dissipate no more than 1 mW in the equilibrium state. The differential voltage gain A_d is to be 10 V/V and the output common-mode dc voltage is to be 0.2 V. (*Note*: This is the dc voltage at the drains.) Assume $\mu_n C_{ox} = 400\ \mu\text{A/V}^2$ and neglect the Early effect. Specify I, R_D, and W/L.

VE 9.1 **D 9.15** Design a MOS differential amplifier to operate from ±0.6-V supplies and dissipate no more than 0.3 mW in its equilibrium state. Select the value of V_{OV} so that the value of v_{id} that steers the current from one side of the pair to the other is 0.2 V. The differential voltage gain A_d is to be 8 V/V. Assume $k_n' = 500\ \mu\text{A/V}^2$ and neglect the Early effect. Specify the required values of I, R_D, and W/L.

9.16 An NMOS differential amplifier with equal drain resistors, $R_D = 47$ kΩ, has a differential gain A_d of 20 V/V.

(a) What is the value of g_m for each of the two transistors?
(b) If each of the two transistors is operating at an overdrive voltage $V_{OV} = 0.2$ V, what must the value of I be?
(c) For $v_{id} = 0$, what is the dc voltage across each R_D?
(d) If v_{id} is 20-mV peak-to-peak sine wave applied in a balanced manner but superimposed on $V_{CM} = 0.5$ V, what is the peak of the sine-wave signal at each drain?
(e) What is the lowest value that V_{DD} must have to ensure saturation-mode operation for Q_1 and Q_2 at all times? Assume $V_t = 0.5$ V.

9.17 A MOS differential amplifier is designed to have a differential gain A_d equal to the voltage gain obtained from a common-source amplifier. Both amplifiers utilize the same values of R_D and supply voltages, and all the transistors have the same W/L ratios. What must the bias current I of the differential pair be relative to the bias current I_D of the CS amplifier? What is the ratio of the power dissipation of the two circuits?

9.18 A differential amplifier is designed to have a differential voltage gain equal to the voltage gain of a common-source amplifier. Both amplifiers use the same values of R_D and supply voltages and are designed to dissipate equal amounts of power in their equilibrium or quiescent state. As well, all the transistors use the same channel length. What must the width W of the differential-pair transistors be relative to the width of the CS transistor?

D 9.19 Figure P9.19 shows a MOS differential amplifer with the drain resistors R_D implemented using diode-connected PMOS transistors, Q_3 and Q_4. Let Q_1 and Q_2 be matched, and Q_3 and Q_4 be matched.

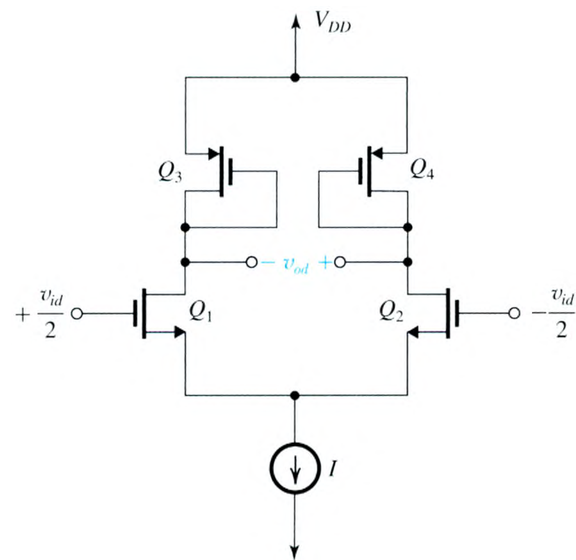


Figure P9.19

(a) Find the differential half-circuit and use it to derive an expression for A_d in terms of $g_{m1,2}$, $g_{m3,4}$, $r_{o1,2}$, and $r_{o3,4}$.
(b) Neglecting the effect of the output resistances r_o, find A_d in terms of μ_n, μ_p, $(W/L)_{1,2}$ and $(W/L)_{3,4}$.
(c) If $\mu_n = 4\mu_p$ and all four transistors have the same channel length, find $(W_{1,2}/W_{3,4})$ that results in $A_d = 10$ V/V.

9.20 Find the differential half-circuit for the differential amplifier shown in Fig. P9.20 and use it to derive an expression for the differential gain $A_d \equiv v_{od}/v_{id}$ in terms of g_m, R_D, and R_s. Neglect the Early effect. What is the gain with $R_s = 0$? What is the value of R_s (in terms of $1/g_m$) that reduces the gain to half this value?

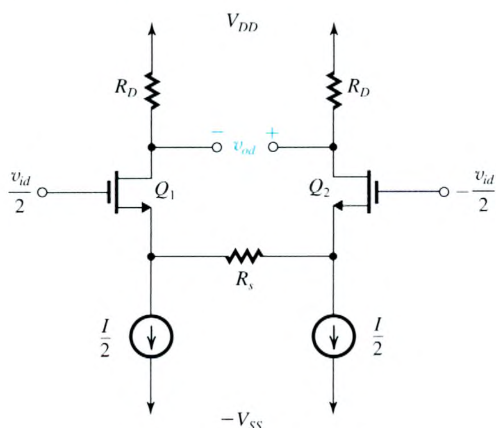


Figure P9.20

***9.21** The resistance R_s in the circuit of Fig. P9.20 can be implemented by using a MOSFET operated in the triode region, as shown in Fig. P9.21. Here Q_3 implements R_s, with the value of R_s determined by the voltage V_C at the gate of Q_3.

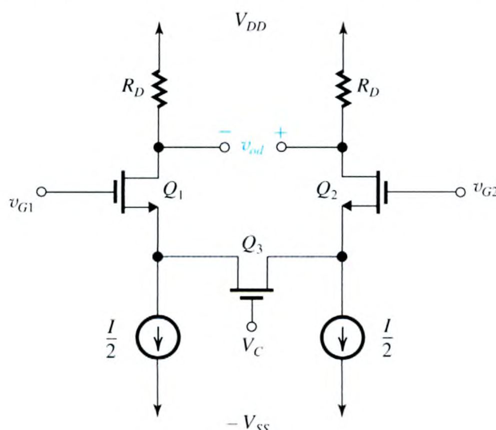


Figure P9.21

(a) With $v_{G1} = v_{G2} = 0$ V, and assuming that Q_1 and Q_2 are operating in saturation, what dc voltages appear at the sources of Q_1 and Q_2? Express these in terms of the overdrive voltage V_{OV} at which each of Q_1 and Q_2 operates, and V_t.
(b) For the situation in (a), what current flows in Q_3? What overdrive voltage V_{OV3} is Q_3 operating at, in terms of V_C, V_{OV}, and V_t?
(c) Now consider the case $v_{G1} = +v_{id}/2$ and $v_{G2} = -v_{id}/2$, where v_{id} is a small signal. Convince yourself that Q_3 now conducts current and operates in the triode region with a small v_{DS}. What resistance r_{DS} does it have, expressed in terms of the overdrive voltage V_{OV3} at which it is operating? This is the resistance R_s. Now if all three transistors have the same W/L, express R_s in terms of V_{OV}, V_{OV3}, and $g_{m1,2}$.
(d) Find V_{OV3} and hence V_C that result in (i) $R_s = 1/g_{m1,2}$; (ii) $R_s = 0.5/g_{m1,2}$.

***9.22** The circuit of Fig. P9.22 shows an effective way of implementing the resistance R_s needed for the circuit in Fig. P9.20. Here R_s is realized as the series equivalent of two MOSFETs Q_3 and Q_4 that are operated in the triode region, thus, $R_s = r_{DS3} + r_{DS4}$. Assume that Q_1 and Q_2 are matched and operate in saturation at an overdrive voltage V_{OV} that corresponds to a drain bias current of $I/2$. Also, assume that Q_3 and Q_4 are matched.

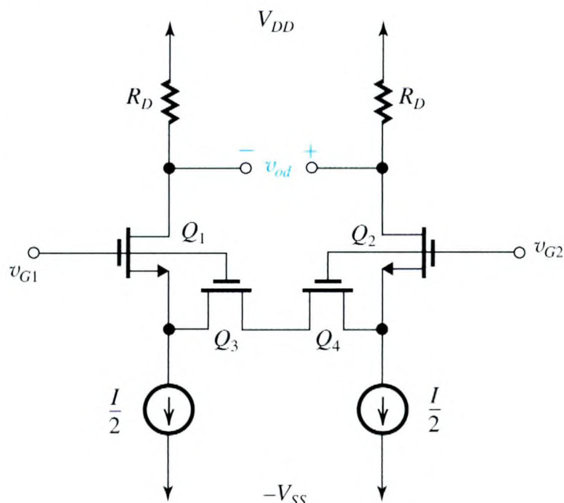


Figure P9.22

(a) With $v_{G1} = v_{G2} = 0$ V, what dc voltages appear at the sources of Q_1 and Q_2? What current flows through Q_3 and Q_4? At what overdrive voltages are Q_3 and Q_4 operating? Find an expression for r_{DS} for each of Q_3 and Q_4 and hence for R_s in terms of $(W/L)_{1,2}$, $(W/L)_{3,4}$, and $g_{m1,2}$.

(b) Now with $v_{G1} = v_{id}/2$ and $v_{G2} = -v_{id}/2$, where v_{id} is a small signal, find an expression of the voltage gain $A_d \equiv v_{od}/v_{id}$ in terms of $g_{m1,2}$, R_D, $(W/L)_{1,2}$, and $(W/L)_{3,4}$.

D *9.23 Figure P9.23 shows a circuit for a differential amplifier with an active load. Here Q_1 and Q_2 form the differential pair, while the current source transistors Q_4 and Q_5 form the active loads for Q_1 and Q_2, respectively. The dc bias circuit that establishes an appropriate dc voltage at the drains of Q_1 and Q_2 is not shown. It is required to design the circuit to meet the following specifications:

(a) Differential gain $A_d = 25$ V/V.
(b) $I_{REF} = I = 200\ \mu$A.
(c) The dc voltage at the gates of Q_6 and Q_3 is +0.4 V.
(d) The dc voltage at the gates of Q_7, Q_4, and Q_5 is −0.4 V.

The technology available is specified as follows: $\mu_n C_{ox} = 4\ \mu_p C_{ox} = 400\ \mu\text{A/V}^2$; $V_{tn} = |V_{tp}| = 0.4$ V, $V_{An} = |V_{Ap}| = 5$ V. Specify the required value of R and the W/L ratios for all transistors. Also specify I_D and $|V_{GS}|$ at which each transistor is operating. For dc bias calculations you may neglect channel-length modulation.

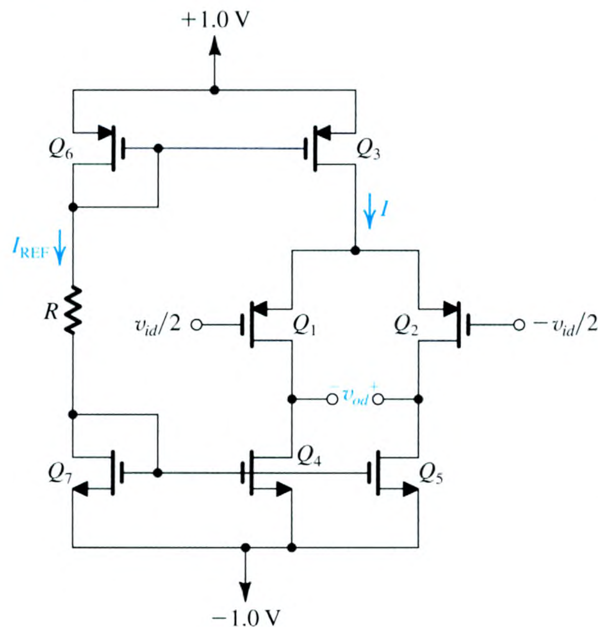


Figure P9.23

***9.24** A design error has resulted in a gross mismatch in the circuit of Fig. P9.24. Specifically, Q_2 has twice the W/L ratio of Q_1. If v_{id} is a small sine-wave signal, find:

(a) I_{D1} and I_{D2}.
(b) V_{OV} for each of Q_1 and Q_2.
(c) The differential gain A_d in terms of R_D, I, and V_{OV}.

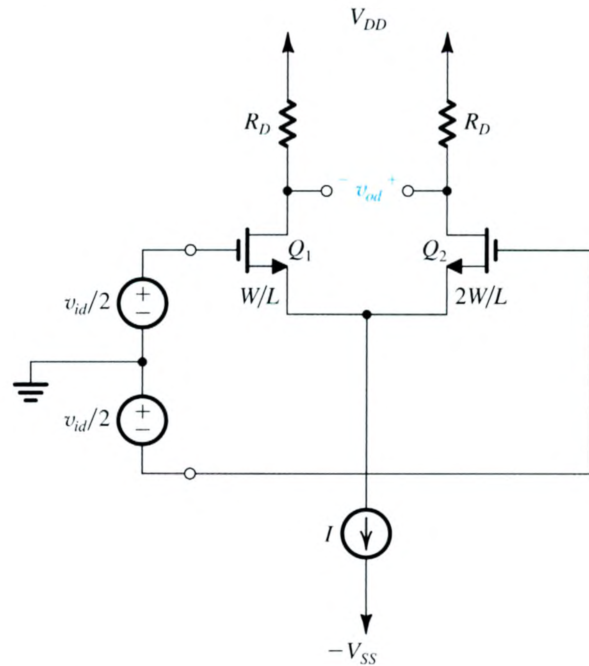


Figure P9.24

D 9.25 For the cascode differential amplifier of Fig. 9.13(a), show that if all transistors have the same channel length and are operated at the same $|V_{OV}|$ and assuming that $V'_{An} = |V'_{Ap}| = |V'_A|$, the differential gain A_d is given by

$$A_d = 2\left(|V_A|/|V_{OV}|\right)^2$$

Now design the amplifier to obtain a differential gain of 600 V/V. Use $|V_{OV}| = 0.15$ V. If $|V'_A| = 5$ V/μm, specify the required channel length L. If g_m is to be as high as possible but the power dissipation in the amplifier (in equilibrium) is to be limited to 0.6 mW, what bias current I would you use? Let $V_{DD} = V_{SS} = 0.9$ V.

Section 9.2: The BJT Differential Pair

9.26 For the differential amplifier of Fig. 9.15(a) let $I = 0.2$ mA, $V_{CC} = V_{EE} = 1.5$ V, $V_{CM} = -0.5$ V, $R_C = 5$ kΩ, and $\beta = 100$. Assume that the BJTs have $v_{BE} = 0.7$ V at $i_C = 1$ mA. Find the voltage at the emitters and at the outputs.

9.27 An *npn* differential amplifier with $I = 0.2$ mA, $V_{CC} = V_{EE} = 1.5$ V, and $R_C = 5$ kΩ utilizes BJTs with $\beta = 100$ and $v_{BE} = 0.7$ V at $i_C = 1$ mA. If $v_{B2} = 0$, find V_E, V_{C1}, and V_{C2} obtained with $v_{B1} = +0.5$ V, and with $v_{B1} = -0.5$ V. Assume that the current source requires a minimum of 0.3 V for proper operation.

9.28 An *npn* differential amplifier with $I = 0.2$ mA, $V_{CC} = V_{EE} = 1.5$ V, and $R_C = 5$ kΩ utilizes BJTs with $\beta = 100$ and $v_{BE} = 0.7$ V at $i_C = 1$ mA. Assuming that the bias current is obtained by a simple current source and that all transistors require a minimum v_{CE} of 0.3 V for operation in the active mode, find the input common-mode range.

9.29 Repeat Exercise 9.7 for an input of –0.3 V.

9.30 An *npn* differential pair employs transistors for which $v_{BE} = 690$ mV at $i_C = 1$ mA, and $\beta = 50$. The transistors leave the active mode at $v_{CE} \leq 0.3$ V. The collector resistors $R_C = 82$ kΩ, and the power supplies are ± 1.2 V. The bias current $I = 20$ μA and is supplied with a simple current source.

(a) For $v_{B1} = v_{B2} = V_{CM} = 0$ V, find V_E, V_{C1}, and V_{C2}.
(b) Find the input common-mode range.
(c) If $v_{B2} = 0$, find the value of v_{B1} that increases the current in Q_1 by 10%.

9.31 Consider the BJT differential amplifier when fed with a common-mode voltage V_{CM} as shown in Fig. 9.15(a). As is often the case, the supply voltage V_{CC} may not be pure dc but might include a ripple component v_r of small amplitude and a frequency of 120 Hz (see Section 4.6). Thus the supply voltage becomes $V_{CC} + v_r$. Find the ripple component of the collector voltages, v_{C1} and v_{C2}, as well as of the difference output voltage $v_{od} \equiv v_{C2} - v_{C1}$. Comment on the differential amplifier response to this undesirable power-supply ripple.

D 9.32 Consider the differential amplifier of Fig. 9.14 and let the BJT β be very large:

(a) What is the largest input common-mode signal that can be applied while the BJTs remain comfortably in the active region with $v_{CB} = 0$?
(b) If the available power supply V_{CC} is 1.5 V, what value of IR_C should you choose in order to allow a common-mode input signal of ± 0.75 V?
(c) For the value of IR_C found in (b), select values for I and R_C. Use the largest possible value for I subject to the constraint that the base current of each transistor (when I divides equally) should not exceed 1 μA. Let $\beta = 100$.

9.33 To gain insight into the possibility of nonlinear distortion resulting from large differential input signals applied to the differential amplifier of Fig. 9.14, evaluate the normalized change in the current i_{E1}, $\Delta i_{E1}/I = \left(i_{E1} - (I/2)\right)/I$, for differential input signals v_{id} of 2, 5, 8, 10, 20, 30, and 40 mV. Use a table to show the ratio $\left(\Delta i_{E1}/I\right)/v_{id}$, which represents the proportional transconductance gain of the differential pair, versus v_{id}. Comment on the linearity of the differential pair as an amplifier.

D 9.34 Design the circuit of Fig. 9.14 to provide a differential output voltage v_O of 1 V when the differential input signal is 10 mV. A current source of 1 mA and a positive supply of +5 V are available. What is the largest possible input common-mode voltage for the circuit to operate as required? Assume $\alpha \simeq 1$.

***9.35** For the circuit in Fig. 9.14, assuming $\alpha = 1$ and $IR_C = 5$ V, use Eqs. (9.48) and (9.49) to find i_{C1} and i_{C2}, and hence determine $v_{od} = v_{C2} - v_{C1}$ for input differential signals $v_{id} \equiv v_{B1} - v_{B2}$ of 2 mV, 5 mV, 10 mV, 15 mV, 20 mV, 25 mV, 30 mV, 35 mV, and 40 mV. Plot v_{od} versus v_{id}, and hence comment on the amplifier linearity. As another way of visualizing linearity, determine the gain $\left(v_o/v_{id}\right)$ versus v_{id}. Comment on the resulting graph.

9.36 In a differential amplifier using a 0.6-mA emitter bias current source, the two BJTs are not matched. Rather, one has twice the emitter junction area of the other. For a differential input signal of zero volts, what do the collector currents become? What difference input is needed to equalize the collector currents? Assume $\alpha = 1$.

***9.37** This problem explores the linearization of the transfer characteristics of the differential pair achieved by including emitter-degeneration resistances R_e in the emitters (see Fig. 9.17). Consider the case $I = 200$ μA with the transistors exhibiting $v_{BE} = 690$ mV at $i_C = 1$ mA and assume $\alpha \simeq 1$.

(a) With no emitter resistances R_e, what value of V_{BE} results when $v_{id} = 0$?
(b) With no emitter resistances R_e, use the large-signal model to find i_{C1} and i_{C2} when $v_{id} = 20$ mV.
(c) Now find the value of R_e that will result in the same i_{C1} and i_{C2} as in (b) but with $v_{id} = 200$ mV. Use the large-signal model.
(d) Calculate the effective transconductance G_m as the ratio of the difference current, $(i_{C1} - i_{C2})$, to v_{id} in the cases without and with the R_e's. By what factor is G_m reduced? How does this factor relate to the increase in v_{id}? Comment.

9.38 A BJT differential amplifier uses a 200-μA bias current. What is the value of g_m of each device? If β is 160, what is the differential input resistance?

D 9.39 Design the basic BJT differential amplifier circuit of Fig. 9.18 to provide a differential input resistance of at least 50 kΩ and a differential voltage gain of 80 V/V. The transistor β is specified to be at least 100. Specify I and R_C.

9.40 For a differential amplifier to which a total difference signal of 10 mV is applied, what is the equivalent signal to its corresponding CE half-circuit? If the emitter current source I is 200 μA, what is r_e of the half-circuit? For a load resistance of 5 kΩ in each collector, what is the half-circuit gain? What magnitude of signal output voltage would you expect at each collector? Between the two collectors?

9.41 A BJT differential amplifier is biased from a 0.5-mA constant-current source and includes a 400-Ω resistor in each emitter. The collectors are connected to V_{CC} via 10-kΩ resistors. A differential input signal of 0.2 V is applied between the two bases.

(a) Find the signal current in the emitters (i_e) and the signal voltage v_{be} for each BJT.
(b) What is the total emitter current in each BJT?
(c) What is the signal voltage at each collector? Assume $\alpha = 1$.
(d) What is the voltage gain realized when the output is taken between the two collectors?


VE 9.2

D 9.42 Design a BJT differential amplifier to amplify a differential input signal of 0.2 V and provide a differential output signal of 2 V. To ensure adequate linearity, it is required to limit the signal amplitude across each base–emitter junction to a maximum of 5 mV. Another design requirement is that the differential input resistance be at least 400 kΩ. The BJTs available are specified to have $\beta \geq 100$. Give the circuit configuration and specify the values of all its components.

D 9.43 Design a bipolar differential amplifier such as that in Fig. 9.18 to operate from ±2.5 V power supplies and to provide differential gain of 60 V/V. The power dissipation in the quiescent state should not exceed 1 mW.

(a) Specify the values of I and R_C. What dc voltage appears at the collectors?
(b) If $\beta = 100$, what is the input differential resistance?
(c) For $v_{id} = 10$ mV, what is the signal voltage at each of the collectors?
(d) For the situation in (c), what is the maximum allowable value of the input common-mode voltage, V_{CM}? Recall that to maintain an *npn* BJT in saturation, v_B should not exceed v_C by more than 0.4 V.

D *9.44 In this problem we explore the trade-off between input common-mode range and differential gain in the design of the bipolar differential BJT amplifier. Consider the bipolar differential amplifier in Fig. 9.14 with the input voltages

$$v_{B1} = V_{CM} + (v_{id}/2)$$
$$v_{B2} = V_{CM} - (v_{id}/2)$$

(a) Bearing in mind that for a BJT to remain in the active mode, v_{BC} should not exceed 0.4 V, show that when v_{id} has a peak $\hat{v}_{id}$, the maximum input common-mode voltage $V_{CM\max}$ is given by

$$V_{CM\max} = V_{CC} + 0.4 - \frac{\hat{v}_{id}}{2} - A_d\left(V_T + \frac{\hat{v}_{id}}{2}\right)$$

(b) For the case $V_{CC} = 2.5$ V and $\hat{v}_{id} = 10$ mV, use the relationship above to determine $V_{CM\max}$ for the case $A_d = 50$ V/V. Also find the peak output signal $\hat{v}_{od}$ and the required value of IR_C. Now if the power dissipation in the circuit is to be limited to 1 mW in the quiescent state (i.e., with $v_{id} = 0$), find I and R_C. (Remember to include the power drawn from the negative power supply $-V_{EE} = -2.5$ V.)
(c) If $V_{CM\max}$ is to be +1 V, and all other conditions remain the same, what maximum gain A_d is achievable?

9.45 For the differential amplifier of Fig. 9.14, let $V_{CC} = +5$ V and $IR_C = 4$ V. Find the differential gain A_d. Sketch and clearly label the waveforms for the total collector voltages v_{C1} and v_{C2} and for (v_O) for the case:

$$v_{B1} = 1 + 0.005 \sin(\omega t)$$
$$v_{B2} = 1 - 0.005 \sin(\omega t)$$

9.46 Consider a bipolar differential amplifier in which the collector resistors R_C are replaced with simple current sources implemented using *pnp* transistors. Sketch the circuit and give its differential half-circuit. If $V_A = 25$ V for all transistors, find the differential voltage gain achieved.

9.47 For each of the emitter-degenerated differential amplifiers shown in Fig. P9.47, find the differential half-circuit and derive expressions for the differential gain A_d and differential input resistance R_{id}. For each circuit, what dc voltage appears across the bias current source(s) in the quiescent state (i.e., with $v_{id} = 0$)? Hence, which of the two circuits will allow a larger negative V_{CM}?

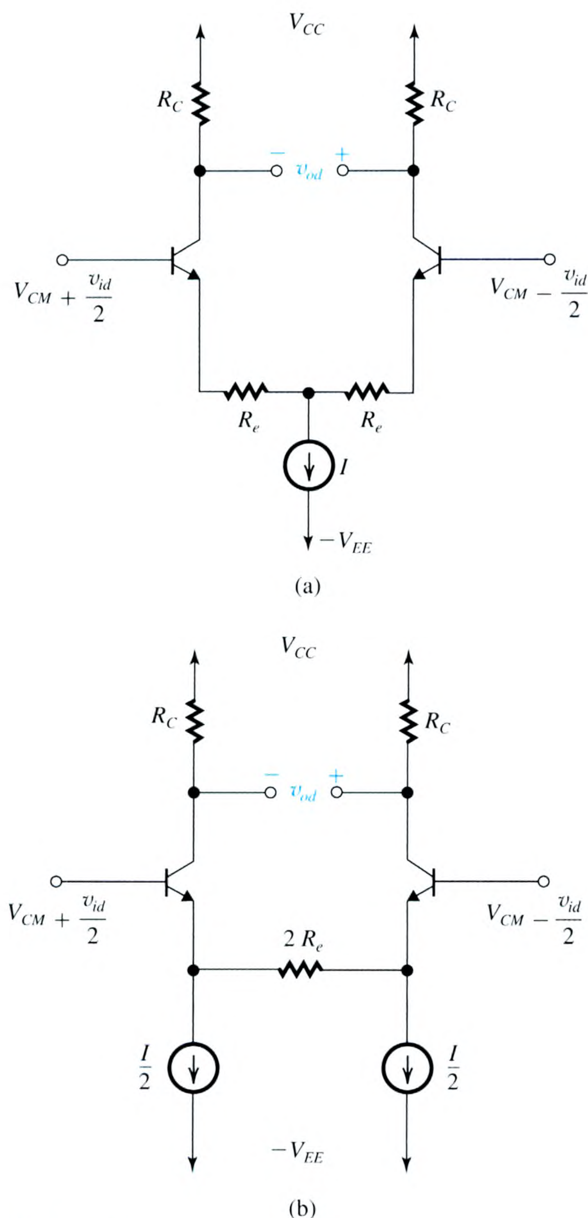


Figure P9.47

9.48 Consider a bipolar differential amplifier that, in addition to the collector resistances R_C, has a load resistance R_L connected between the two collectors. What does the differential gain A_d become?

9.49 A bipolar differential amplifier having resistance R_e inserted in series with each emitter [as in Fig. P9.47(a)] is biased with a constant current I. When both input terminals are grounded, the dc voltage measured across each R_e is found to be 3 V_T and that measured across each R_C is found to be $80V_T$. What differential voltage gain A_d do you expect the amplifier to have?

9.50 A bipolar differential amplifier with emitter-degeneration resistances R_e and R_e is fed with the arrangement shown in Fig. P9.50. Derive an expression for the overall differential voltage gain $G_v \equiv v_{od}/v_{sig}$. If R_{sig} is of such a value that $v_{id} = 0.5v_{sig}$, find the gain G_v in terms of R_C, r_e, R_e, and α. Now if β is doubled, by what factor does G_v increase?

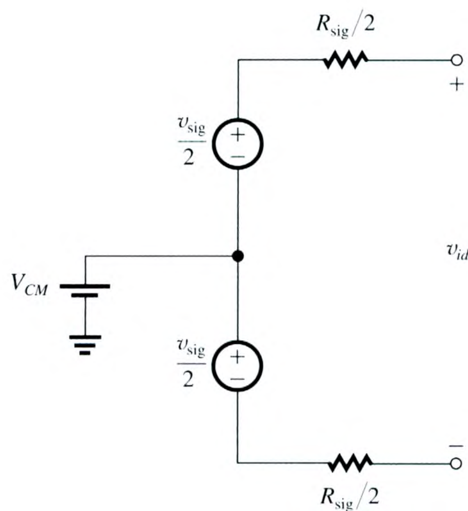


Figure P9.50

9.51 A particular differential amplifier operates from an emitter current source $I = 0.2$ mA. Each of the collector resistances $R_C = 20$ kΩ and a load resistance $R_L = 40$ kΩ is connected between the two collectors. If the amplifier is fed in the manner shown in Fig. P9.50 with $R_{sig} = 80$ kΩ, find the overall voltage gain. Let $\beta = 160$.

9.52 Find the voltage gain and the input resistance of the amplifier shown in Fig. P9.52 assuming $\beta = 100$.

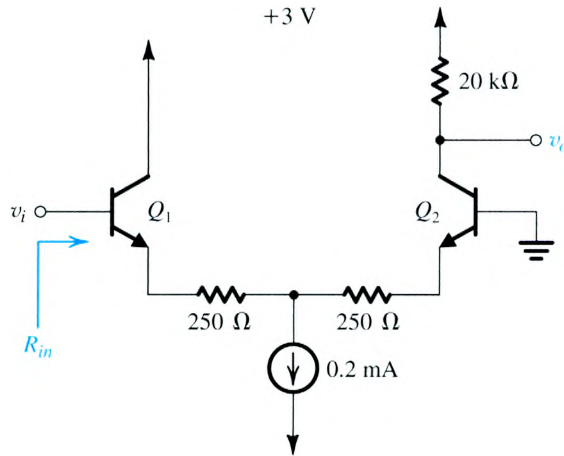


Figure P9.52

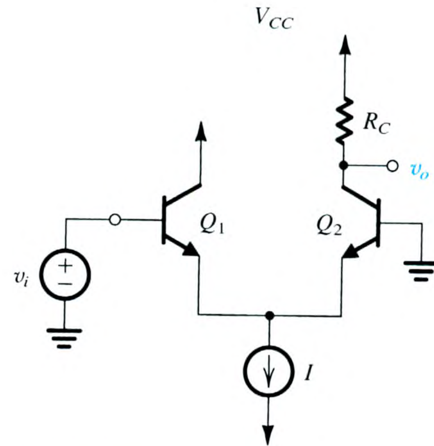


Figure P9.54

9.53 Find the voltage gain and input resistance of the amplifier in Fig. P9.53 assuming that $\beta = 100$.

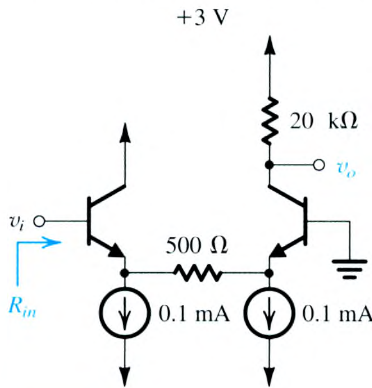


Figure P9.53

9.54 Derive an expression for the small-signal voltage gain v_o/v_i of the circuit shown in Fig. P9.54 in two different ways:

(a) as a differential amplifier
(b) as a cascade of a common-collector stage Q_1 and a common-base stage Q_2

Assume that the BJTs are matched and have a current gain α, and neglect the Early effect. Verify that both approaches lead to the same result.

Section 9.3: Common-Mode Rejection

VE 9.3 SIM **9.55** An NMOS differential pair is biased by a current source $I = 0.2$ mA having an output resistance $R_{SS} = 100$ kΩ. The amplifier has drain resistances $R_D = 12$ kΩ, using transistors with $k_n'W/L = 5$ mA/V^2, and r_o that is large. If the output is taken differentially and there is a 1% mismatch between the drain resistances, find $|A_d|$, $|A_{cm}|$, and CMRR.

9.56 For the differential amplifier shown in Fig. P9.2, let Q_1 and Q_2 have $k_p'(W/L) = 5$ mA/V^2, and assume that the bias current source has an output resistance of 50 kΩ. Find $|V_{OV}|$, g_m, $|A_d|$, $|A_{cm}|$, and the CMRR (in dB) obtained with the output taken differentially. The drain resistances are known to have a mismatch of 2%.

SIM **D *9.57** The differential amplifier in Fig. P9.57 utilizes a resistor R_{SS} to establish a 0.02-mA dc bias current. Note that this amplifier uses a single 2-V supply and thus the dc common-mode voltage V_{CM} cannot be zero. Transistors Q_1 and Q_2 have $k_n'W/L = 1$ mA/V^2, $V_t = 0.4$ V, and $\lambda = 0$.

(a) Find the required value of V_{CM}.
(b) Find the value of R_D that results in a differential gain A_d of 15 V/V.
(c) Determine the dc voltage at the drains.

problems with blue numbers are considered essential; * = difficult problem; ** = more difficult; *** = very challenging
SIM = simulation; **D** = design problem; ● = see related video example

(d) Determine the single-ended-output common-mode gain $\Delta V_{D1}/\Delta V_{CM}$. (*Hint:* You need to take $1/g_m$ into account.)
(e) Use the common-mode gain found in (d) to find the change in V_{CM} that results in Q_1 and Q_2 entering the triode region.

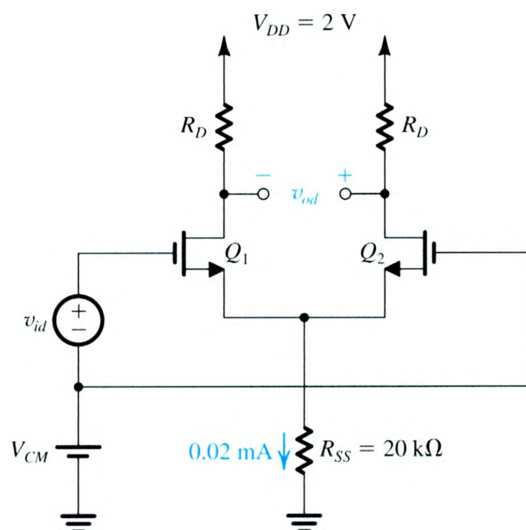


Figure P9.57

9.58 It can be shown that if the drain resistors of a MOS differential amplifier have a mismatch ΔR_D and if simultaneously the transconductances of Q_1 and Q_2 have a mismatch Δg_m, the common-mode gain is given by

$$A_{cm} \simeq \left(\frac{R_D}{2R_{SS}}\right)\left(\frac{\Delta g_m}{g_m} + \frac{\Delta R_D}{R_D}\right)$$

Note that this equation indicates that R_D can be deliberately varied to compensate for the initial variability in g_m and R_D, that is, to minimize A_{cm}.

In a MOS differential amplifier for which $R_D = 5$ kΩ and $R_{SS} = 50$ kΩ, the common-mode gain is found to be 0.002 V/V. Find the percentage change required in one of the two drain resistors to reduce A_{cm} to zero (or close to zero).

D 9.59 Design a MOS differential amplifier to have a CMRR of 74 dB. The only source of mismatch in the circuit is a 1% difference between the *W/L* ratios of the two transistors. Let $I = 100$ µA and assume that all transistors are operated at $V_{OV} = 0.2$ V. For the 0.18-µm CMOS fabrication process available, $V_A' = 5$ V/µm. What is the value of L required for the current-source transistor? If the current-source transistor is cascoded with an idential transistor, what will the CMRR value become?

D 9.60 A MOS differential amplifier utilizing a simple current source to provide the bias current I has a CMRR of 60 dB. If the CMRR must be raised to 100 dB by adding a cascode transistor to the current source, what must the intrinsic gain A_0 of the cascode transistor be? If the cascode transistor is operated at $V_{OV} = 0.2$ V, what must its V_A be? If for the specific technology utilized $V_A' = 5$ V/µm, specify the channel length L of the cascode transistor.

9.61 The differential amplifier circuit of Fig. P9.61 utilizes a resistor connected to the negative power supply to establish the bias current I.

(a) For $v_{B1} = v_{id}/2$ and $v_{B2} = -v_{id}/2$, where v_{id} is a small signal with zero average, find the magnitude of the differential gain, $|v_o/v_{id}|$.
(b) For $v_{B1} = v_{B2} = v_{icm}$, where v_{icm} has a zero average, find the magnitude of the common-mode gain, $|v_o/v_{icm}|$.
(c) Calculate the CMRR.
(d) If $v_{B1} = 0.1 \sin 2\pi \times 60t + 0.005 \sin 2\pi \times 1000t$, volts, and $v_{B2} = 0.1 \sin 2\pi \times 60t - 0.005 \sin 2\pi \times 1000t$, volts, find v_o.

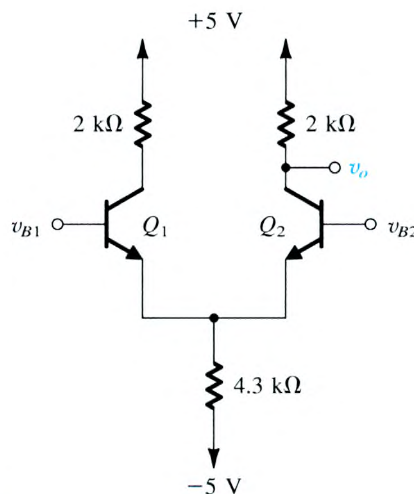


Figure P9.61

9.62 Consider the basic differential circuit in which the transistors have $\beta = 120$ and $V_A = 80$ V, with $I = 0.2$ mA, $R_{EE} = 400$ kΩ, and $R_C = 20$ kΩ. The collector resistances are matched to within 2%. Find:

(a) the differential gain
(b) the differential input resistance
(c) the common-mode gain

(d) the common-mode rejection ratio
(e) the common-mode input resistance

9.63 For the differential amplifier shown in Fig. P9.63, identify and sketch the differential half-circuit and the common-mode half-circuit. Find the differential gain, the differential input resistance, the common-mode gain assuming the resistances R_C have 1% tolerance, and the common-mode input resistance. For these transistors, $\beta = 100$ and $V_A = 100$ V.

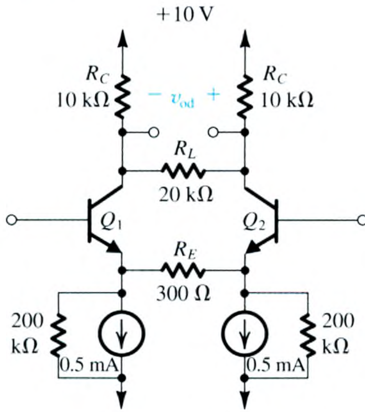


Figure P9.63

9.64 In a bipolar differential-amplifier circuit, the bias current generator consists of a simple common-emitter transistor operating at 200 μA. For this transistor, and those used in the differential pair, $V_A = 30$ V and $\beta = 60$. What common-mode input resistance would result? Assume $R_C \ll r_o$.

9.65 A bipolar differential amplifier with $I = 0.2$ mA utilizes transistors for which $V_A = 20$ V and $\beta = 100$. The collector resistances $R_C = 10$ kΩ and are matched to within 10%. Find:

(a) the differential gain
(b) the common-mode gain and the CMRR if the bias current I is generated using a simple current mirror
(c) the common-mode gain and the CMRR if the bias current I is generated using a Wilson mirror. (Refer to Eq. 8.98 for R_o of the Wilson mirror.)

D 9.66 Design a differential amplifier to provide the largest possible signal to a pair of 10-kΩ load resistances. The input differential signal is a sinusoid of 5-mV peak amplitude, which is applied to one input terminal while the other input terminal is grounded. The power supply V_{CC} available is 5 V. To determine the required bias current I, derive an expression for the total voltage at each of the collectors in terms of V_{CC} and I in the presence of the input signal. Then impose the condition that both transistors remain well out of saturation with a minimum v_{CB} of roughly 0 V. Thus determine the required value of I. For this design, what differential gain is achieved? What is the amplitude of the signal voltage obtained between the two collectors? Assume $\alpha \simeq 1$.

D *9.67 Design a BJT differential amplifier that provides two single-ended outputs (at the collectors). The amplifier is to have a differential gain (to each of the two outputs) of at least 100 V/V, a differential input resistance ≥ 10 kΩ, and a common-mode gain (to each of the two outputs) no greater than 0.1 V/V. Use a 2-mA current source for biasing. Give the complete circuit with component values and suitable power supplies that allow for ±2 V swing at each collector and an input common-mode voltage as high as +3 V. Specify the minimum value that the output resistance of the bias current source must have. If the current source is realized by a simple mirror, what must the minimum value of V_A be? The BJTs available have $\beta \geq 100$. What is the value of the input common-mode resistance when the bias source has the lowest acceptable output resistance?

9.68 When the output of a BJT differential amplifier is taken differentially, its CMRR is found to be 46 dB higher than when the output is taken single-endedly. If the only source of common-mode gain when the output is taken differentially is the mismatch in collector resistances, what must this mismatch be (in percent)?

***9.69** In a particular BJT differential amplifier, a production error results in one of the transistors having an emitter–base junction area that is twice that of the other. With the inputs grounded, how will the emitter bias current split between the two transistors? If the output resistance of the current source is 400 kΩ and the resistance in each collector (R_C) is 15 kΩ, find the common-mode gain obtained when the output is taken differentially. Assume $\alpha \simeq 1$. (*Hint*: The CM signal current v_{icm}/R_{EE} will split between Q_1 and Q_2 in the same ratio as the bias current I does.)

Section 9.4: DC Offset

D 9.70 An NMOS differential pair is to be used in an amplifier whose drain resistors are 10 kΩ ± 1%. For the pair, $k_n' W/L = 4$ mA/V^2. Decide whether to use a bias current I of 160 μA or 360 μA. Contrast the differential gain and input offset voltage for the two possibilities.

D 9.71 An NMOS differential amplifier for which the MOSFETs have a transconductance parameter k_n and whose drain resistances R_D have a mismatch ΔR_D is biased with a current I.

(a) Find expressions for A_d and V_{OS} in terms of k_n, R_D, $\Delta R_D/R_D$, and I. Use these expressions to relate V_{OS} and A_d.

(b) If $k_n = 4$ mA/V^2, $R_D = 10$ kΩ, and $\Delta R_D/R_D = 0.02$, find the maximum gain realized if V_{OS} is to be limited to 1 mV, 2 mV, 3 mV, 4 mV, and 5 mV. For each case, give the value of the required bias current I. Note the trade-off between gain and offset voltage.

D 9.72 An NMOS amplifier, whose designed operating point is at $V_{OV} = 0.2$ V, is suspected to have a variability of V_t of ±4 mV, and of W/L and R_D (independently) of ±1%. What is the worst-case input offset voltage you would expect to find? What is the major contribution to this total offset? If you used a variation of one of the drain resistors to reduce the output offset to zero and thereby compensate for the uncertainties (including that of the other R_D), what percentage change from nominal would you require?

9.73 An NMOS differential pair operating at a bias current I of 100 μA uses transistors for which $k_n' = 400$ μA/V^2 and $W/L = 10$. Find the three components of input offset voltage under the conditions that $\Delta R_D/R_D = 4\%$, $\Delta(W/L)/(W/L) = 4\%$, and $\Delta V_t = 5$ mV. In the worst case, what might the total offset be? For the usual case of the three effects being independent, what is the offset likely to be?

9.74 A bipolar differential amplifier uses two well-matched transistors, but collector load resistors that are mismatched by 8%. What input offset voltage is required to reduce the differential dc output voltage to zero?

9.75 A bipolar differential amplifier uses two transistors whose scale currents I_S differ by 8%. If the two collector resistors are well matched, find the resulting input offset voltage.

9.76 Modify Eq. (9.101) for a differential amplifier having a resistance R_E connected in the emitter of each transistor. Let the bias current source be I.

9.77 A differential amplifier uses two transistors whose β values are β_1 and β_2. If everything else is matched, show that the input offset voltage is approximately $V_T[(1/\beta_1) - (1/\beta_2)]$. Evaluate V_{OS} for $\beta_1 = 50$ and $\beta_2 = 100$. (*Hint:* The bias current I will still split equally between the two emitters.)

9.78 Two possible differential amplifier designs are considered, one using BJTs and the other MOSFETs. In both cases, the collector (drain) resistors are maintained within ±2% of nominal value. The MOSFETs are operated at $V_{OV} = 200$ mV. What input offset voltage results in each case? What does the MOS V_{OS} become if the devices are increased in width by a factor of 4 while the bias current is kept constant?

***9.79** A differential amplifier uses two transistors having V_A values of 100 V and 200 V. If everything else is matched, find the resulting input offset voltage. Assume that the two transistors are intended to be biased at a V_{CE} of about 10 V.

****9.80** A differential amplifier is fed in a balanced or push–pull manner, and the source resistance in series with each base is R_s. Show that a mismatch ΔR_s between the values of the two source resistances gives rise to an input offset voltage of approximately $(I/2\beta)\Delta R_s / [1 + (g_m R_s)/\beta]$.

9.81 One approach to "offset correction" involves adjusting the values of R_{C1} and R_{C2} to reduce the differential output voltage to zero when both input terminals are grounded. This offset-nulling process can be accomplished by using a potentiometer in the collector circuit, as shown in Fig. P9.81. Find the potentiometer setting, represented by the fraction x of its value connected in series with R_{C1}, that is required for nulling the output offset voltage that results from:

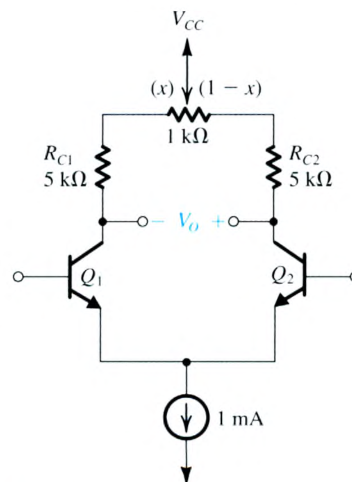


Figure P9.81

(a) R_{C1} being 5% higher than nominal and R_{C2} 5% lower than nominal.
(b) Q_1 having an area 4% larger than nominal, while Q_2 has area 4% smaller than nominal.

9.82 A differential amplifier for which the total emitter bias current is 500 μA uses transistors for which β is specified to lie between 80 and 200. What is the largest possible input bias current? The smallest possible input bias current? The largest possible input offset current?

***9.83** In a particular BJT differential amplifier, a production error results in one of the transistors having an emitter–base junction area twice that of the other. With both inputs grounded, find the current in each of the two transistors and hence the dc offset voltage at the output, assuming that the collector resistances are equal. Use small-signal analysis to find the input voltage that would restore current balance to the differential pair. Repeat using large-signal analysis and compare results.

D 9.84 A large fraction of mass-produced differential-amplifier modules using 20-kΩ collector resistors is found to have an input offset voltage ranging from +2 mV to –2 mV. By what amount must one collector resistor be adjusted to reduce the input offset to zero? If an adjustment mechanism is devised that raises one collector resistance while correspondingly lowering the other, what resistance change is needed? If a potentiometer connected as shown in Fig. P9.81 is used, what value of potentiometer resistance (specified to 1 significant digit) is needed? Assume that the offset is entirely due to the finite tolerance of R_C.

Section 9.5: The Differential Amplifier with a Current-Mirror Load

9.85 The differential amplifier of Fig. 9.31(a) is found to have a differential short-circuit transconductance of 2 mA/V. A differential input signal is applied and the output voltage is measured with a load resistance R_L connected. It is found that when R_L is reduced from ∞ to 25 kΩ, the magnitude of the output signal is reduced by half. What do you estimate R_o and A_d (with R_L disconnected) to be?

9.86 A current-mirror-loaded NMOS differential amplifier is fabricated in a technology for which $|V_A'| = 6$ V/μm. All the transistors have $L = 0.5$ μm. If the differential-pair transistors are operated at $V_{OV} = 0.2$ V, what open-circuit differential gain is realized?

9.87 The differential amplifier of Fig. 9.31(a) is biased with $I = 200$ μA. All transistors have $L = 0.5$ μm, and Q_1 and Q_2 have $W/L = 20$. The circuit is fabricated in a process for which $\mu_n C_{ox} = 400$ μA/V^2 and $|V_A'| = 6$ V/μm. Find $g_{m1,2}$, r_{o2}, r_{o4}, and A_d.

D 9.88 In a current-mirror-loaded differential amplifier of the form shown in Fig. 9.31(a), all transistors are characterized by $k'W/L = 4$ mA/V^2, and $|V_A| = 5$ V. Find the bias current I for which the gain $v_o/v_{id} = 25$ V/V.

D 9.89 Consider a current-mirror-loaded differential amplifier such as that shown in Fig. 9.31(a) with the bias current source implemented with the modified Wilson mirror of Fig. P9.89 with $I = 200$ μA. The transistors have $|V_t| = 0.4$V and $k'W/L = 5$ mA/V^2. What is the lowest value of the total power supply $(V_{DD} + V_{SS})$ that allows each transistor to operate with $|V_{DS}| \geq |V_{GS}|$?

***9.90** (a) Sketch the circuit of a current-mirror-loaded MOS differential amplifier in which the input transistors are cascoded and a cascode current mirror is used for the load.
(b) Show that if all transistors are operated at an overdrive voltage V_{OV} and have equal Early voltages $|V_A|$, the gain is given by

$$A_d = 2(V_A/V_{OV})^2$$

Evaluate the gain for $V_{OV} = 0.20$ V and $V_A = 10$ V.

9.91 Figure P9.91 shows the current-mirror-loaded MOS differential amplifier prepared for small-signal analysis. We have pulled r_o out of each transistor; thus, the current in the drain of each transistor will be $g_m v_{gs}$. We have also simplified matters by indicating approximate values for some of the node voltages. For instance, the output voltage $v_o = \frac{1}{2}(g_m r_o)v_{id}$, which we have derived in the text. The voltage at the common sources has been found to be roughly $+v_{id}/4$, which is very far from the virtual ground we might expect. Also, the voltage at the drain of Q_1 is approximately $-v_{id}/4$, confirming our contention that the voltage there is vastly different from the output voltage, hence the lack of balance in the circuit and the unavailability of a differential half-circuit. Find the currents labeled i_1 to i_{13} in terms of

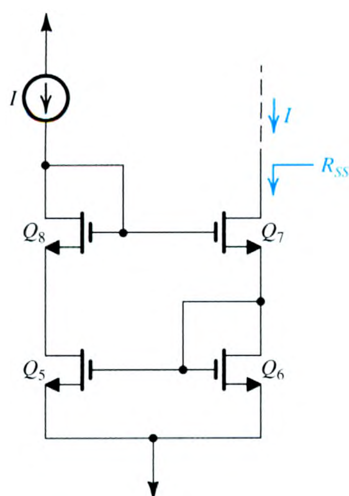


Figure P9.89

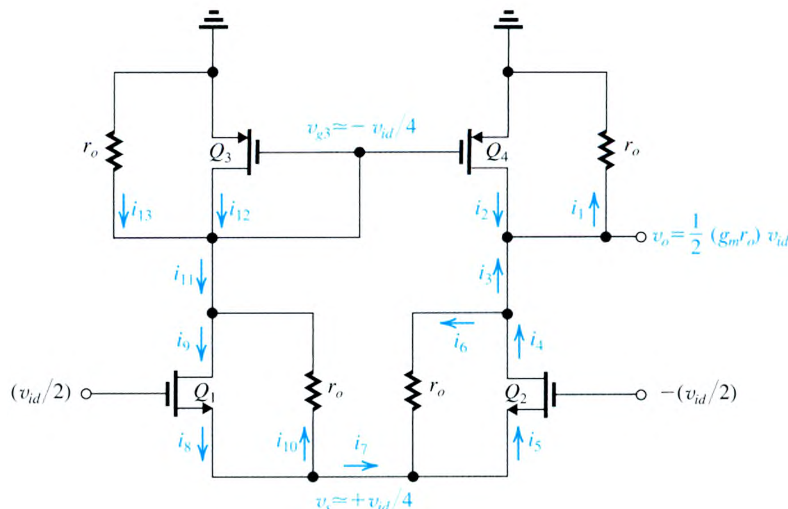


Figure P9.91

$(g_m v_{id})$. Determine their values in the sequence of their numbering and assume $g_m r_o \gg 1$. Assume that all transistors are operating at the same $|V_{OV}|$. Write the current values on the circuit diagram and reflect on the results.

9.92 A current-mirror-loaded NMOS differential amplifier operates with a bias current I of 200 μA. The NMOS transistors are operated at $V_{OV} = 0.2$ V and the PMOS devices at $|V_{OV}| = 0.3$ V. The Early voltages are 10 V for the NMOS and 8 V for the PMOS transistors. Find G_{md}, R_o, and A_d. For what value of load resistance is the gain reduced by a factor of 2?

9.93 This problem investigates the effect of transistor mismatches on the input offset voltage of the current-mirror-loaded MOS differential amplifier of Fig. 9.31(a). For this purpose, ground both input terminals and short-circuit the output node to ground.

(a) If the amplifying transistors Q_1 and Q_2 exhibit a W/L mismatch of $\Delta(W/L)_A$, find the resulting short-circuit output current and hence show that the corresponding V_{OS} is given by

$$V_{OS1} = (V_{OV}/2)\frac{\Delta(W/L)_A}{(W/L)_A}$$

where V_{OV} is the overdrive voltage at which Q_1 and Q_2 are operating.

(b) Repeat for a mismatch $\Delta(W/L)_M$ in the W/L ratios of the mirror transistor Q_3 and Q_4 to show that the corresponding V_{OS} is given by

$$V_{OS2} = (V_{OV}/2)\frac{\Delta(W/L)_M}{(W/L)_M}$$

where V_{OV} is the overdrive voltage at which Q_1 and Q_2 are operating.

(c) For a circuit in which all transistors are operated at $|V_{OV}| = 0.2$ V and all W/L ratios are accurate to within ±1% of nominal, find the worst-case total offset voltage V_{OS}.

9.94 The differential amplifier in Fig. 9.35(a) is operated with $I = 400$ μA, with devices for which $V_A = 10$ V and $\beta = 100$. What differential input resistance, output resistance, differential short-circuit transconductance, and open-circuit voltage gain would you expect? What will the voltage gain be if the input resistance of the subsequent stage is equal to R_{id} of this stage?

9.95 Figure P9.95 shows a differential cascode amplifier with an active load formed by a Wilson current mirror. Utilizing the expressions derived in Chapter 8 for the output resistance of a bipolar cascode and the output resistance of the Wilson mirror, and assuming all transistors to be identical, show that the differential voltage gain A_d is given approximately by

$$A_d = \frac{1}{3}\beta g_m r_o$$

Evaluate A_d for the case of $\beta = 100$ and $V_A = 20$ V.

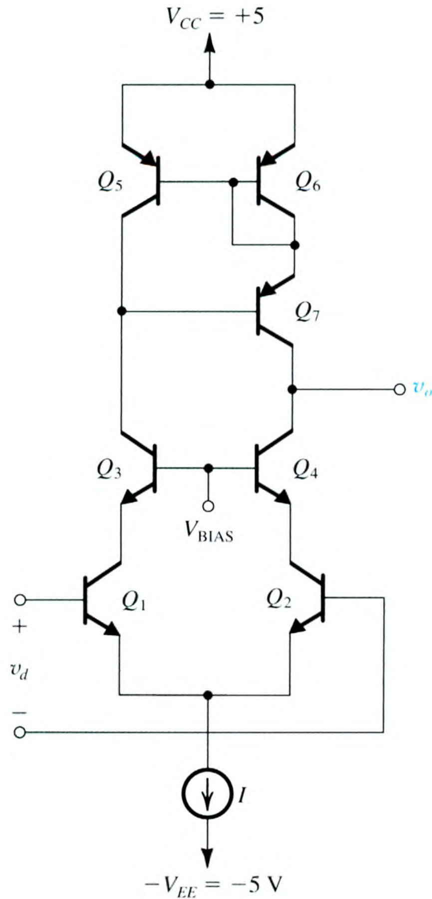


Figure P9.95

D 9.96 Consider the bias design of the Wilson-loaded cascode differential amplifier shown in Fig. P9.95.

(a) What is the largest signal voltage possible at the output without Q_7 saturating? Assume that the CB junction conducts when the voltage across it exceeds 0.4 V.

(b) What should the dc bias voltage established at the output (by an arrangement not shown) be in order to allow for positive output signal swing of 1.5 V?

(c) What should the value of V_{BIAS} be in order to allow for a negative output signal swing of 1.5 V?

(d) What is the upper limit on the input common-mode voltage V_{ICM}?

9.97 In addition to the random offset voltages that result from the mismatches inevitably present in the differential amplifier, the current-mirror-loaded bipolar differential amplifier suffers from a *systematic* offset voltage. This is due to the error in the current transfer ratio of the current-mirror load caused by the finite β of the *pnp* transistors that make up the mirror. The situation is illustrated in Fig. P9.97. Show that the resulting input offset voltage is given by

$$V_{OS} = -2V_T/\beta_P$$

and evaluate V_{OS} for $\beta_P = 50$. Also, if the simple current mirror is replaced by the *pnp* version of the Wilson mirror of Fig. 8.37(a), find the expected systematic input offset voltage. Give both an expression and a value.

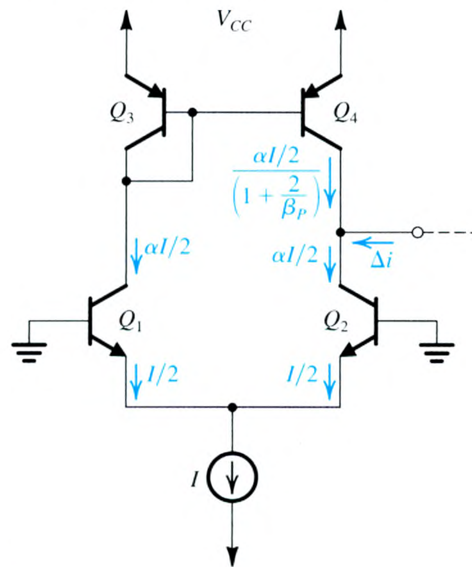


Figure P9.97

9.98 For the BiCMOS differential amplifier in Fig. P9.98 let $V_{DD} = V_{SS} = 3$ V, $I = 0.2$ mA, $k_p'W/L = 6.4$ mA/V^2; $|V_A|$ for p-channel MOSFETs is 10 V, $|V_A|$ for *npn* transistors is 30 V. Find G_{md}, R_o, and A_d.

Figure P9.98

SIM **D 9.99** Design the current-mirror-loaded differential MOS amplifier of Fig. 9.31(a) to obtain a differential gain of 60 V/V. The technology available provides $\mu_n C_{ox} = 4\mu_p C_{ox} = 500\ \mu\text{A/V}^2$, $|V_t| = 0.4$ V, and $|V_A'| = 15$ V/μm and operates from ±0.8-V supplies. Use a bias current $I = 200$ μA and operate all devices at $|V_{OV}| = 0.15$ V. Assume all transistors have the same channel length.

VE 9.4

(a) Find the W/L ratios of the four transistors.
(b) Specify the channel length required of all transistors.
(c) If $V_{ICM} = 0$, what is the allowable range of v_O?
(d) If I is delivered by a simple NMOS current source operated at the same V_{OV} and having the same channel length as the other four transistors, determine the CMRR obtained.

9.100 Consider the current-mirror-loaded MOS differential amplifier of Fig. 9.31(a) in two cases:

(a) Current source I is implemented with a simple current mirror.
(b) Current source I is implemented with the modified Wilson current mirror shown in Fig. P9.100.

Recalling that for the simple mirror $R_{SS} = r_o|_{Q5}$ and for the Wilson mirror $R_{SS} \simeq g_{m7} r_{o7} r_{o5}$, and assuming that all transistors have the same $|V_A|$ and $k'W/L$, show that for (a)

$$\text{CMRR} = 2\left(\frac{V_A}{V_{OV}}\right)^2$$

and for (b)

$$\text{CMRR} = 2\sqrt{2}\left(\frac{V_A}{V_{OV}}\right)^3$$

where V_{OV} is the overdrive voltage that corresponds to a drain current of $I/2$. For $k'W/L = 4\ \text{mA/V}^2$, $I = 160$ μA, and $|V_A| = 5$ V, find CMRR for both cases.

Figure P9.100

9.101 The MOS differential amplifier of Fig. 9.31(a) is biased with a simple current mirror delivering $I = 200$ μA. All transistors are operated at $V_{OV} = 0.2$ V and have $V_A = 6$ V. Find G_{md}, R_o, A_d, R_{SS}, R_{im}, A_m, G_{mcm}, A_{cm}, and CMRR.

9.102 A current-mirror-loaded MOS differential amplifier is found to have a differential voltage gain A_d of 30 V/V. Its bias current source has an output resistance $R_{SS} = 100$ kΩ. The current mirror utilized has a current gain A_m of 0.98 A/A, an input resistance R_{im} of 1.63 kΩ, and an output resistance R_{om} of 75 kΩ. If the output resistances of the amplifier R_{od} is 50 kΩ, find G_{mcm}, A_{cm}, and CMRR.

D 9.103 A current-mirror-loaded MOS differential amplifier has a differential voltage gain A_d of 40 V/V. The output resistance of the current-mirror load is equal to the output resistance of the amplifier at 40 kΩ. If the current-mirror gain is 0.98 A/A and its input resistance R_{im} is 500 Ω, find the required value of the output resistance R_{SS} of the bias current source to obtain a CMRR of 80 dB.

Section 9.6: Multistage Amplifiers

9.104 Consider the circuit in Fig. 9.37 with the device geometries (in μm) shown in Table P9.104. Let $I_{REF} = 50$ μA, $|V_t| = 0.4$ V for all devices, $\mu_n C_{ox} =$

Table P9.104

Transistor	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8
W/L	25/0.5	25/0.5	6.25/0.5	6.25/0.5	50/0.5	W/0.5	50/0.5	12.5/0.5

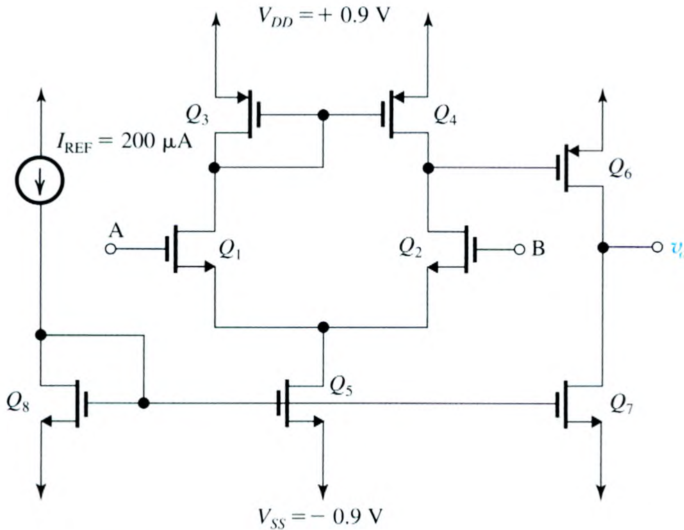


Figure P9.105

360 μA/V^2, $\mu_p C_{ox} = 90$ μA/V^2, $|V_A| = 6$ V for all devices, $V_{DD} = V_{SS} = 0.9$ V. Determine the width of Q_6, W, that will ensure that the op amp will not have a systematic offset voltage. Then, for all devices evaluate I_D, $|V_{OV}|$, $|V_{GS}|$, g_m, and r_o. Provide your results in a table similar to Table 9.1. Also find A_1, A_2, the open-loop voltage gain, the input common-mode range, and the output voltage range. Neglect the effect of V_A on the bias currents.

D 9.105 The two-stage CMOS op amp in Fig. P9.105 is fabricated in a 0.18-μm technology having $k_n' = 4k_p' = 400$ μA/V^2, $V_{tn} = -V_{tp} = 0.4$ V.

(a) With A and B grounded, perform a dc design that will result in each of Q_1, Q_2, Q_3, and Q_4 conducting a drain current of 100 μA and each of Q_6 and Q_7 a current of 200 μA. Design so that all transistors operate at 0.2-V overdrive voltages. Specify the W/L ratio required for each MOSFET. Present your results in tabular form. What is the dc voltage at the output (ideally)?
(b) Find the input common-mode range.
(c) Find the allowable range of the output voltage.
(d) With $v_A = v_{id}/2$ and $v_B = -v_{id}/2$, find the voltage gain v_o/v_{id}. Assume an Early voltage of 6 V.

D *9.106 In a particular design of the CMOS op amp of Fig. 9.37 the designer wishes to investigate the effects of increasing the W/L ratio of both Q_1 and Q_2 by a factor of 4. Assuming that all other parameters are kept unchanged, refer to Example 9.6 to help you answer the following questions:

(a) What change results in $|V_{OV}|$ and in g_m of Q_1 and Q_2?
(b) What change results in the voltage gain of the input stage? In the overall voltage gain?
(c) What is the effect on the input offset voltages? (You might wish to refer to Section 9.4).

9.107 Consider the amplifier of Fig. 9.37, whose parameters are specified in Example 9.6. If a manufacturing error results in the W/L ratio of Q_7 being 24/0.4, find the current that Q_7 will now conduct. Thus find the systematic offset voltage that will appear at the output. (Use the results of Example 9.6.)

Assuming that the open-loop gain will remain approximately unchanged from the value found in Example 9.6, find the corresponding value of input offset voltage, V_{OS}.

9.108 Consider the input stage of the CMOS op amp in Fig. 9.37 with both inputs grounded. Assume that the two sides of the input stage are perfectly matched except that the threshold voltages of Q_3 and Q_4 have a mismatch ΔV_t. Show that a current $g_{m3}\Delta V_t$ appears at the output of the first stage. What is the corresponding input offset voltage?

D 9.109 The two-stage op amp in Figure P9.105 is fabricated in a 65-nm technology having $k_n' = 5.4 \times k_p' = 540\ \mu\text{A/V}^2$ and $V_{tn} = -V_{tp} = 0.35$ V. The amplifier is operated with $V_{DD} = +1.2$ V and $V_{SS} = 0$ V.

(a) Perform a dc design that will cause each of Q_1, Q_2, Q_3, and Q_4 to conduct a drain current of 200 μA and each of Q_6 and Q_7 to conduct a current of 400 μA. Design so that all transistors operate at 0.15-V overdrive voltages. Specify the *W/L* ratio required for each MOSFET. Present all results in a table.
(b) Find the input common-mode range.
(c) Find the allowable range of the output voltage.
(d) With $v_A = v_{id}/2$ and $v_B = -v_{id}/2$, find the voltage gain v_o/v_{id}. Assume an Early voltage of 2.4 V.

***9.110** Figure P9.110 shows a bipolar op-amp circuit that resembles the CMOS op amp of Fig. 9.37. Here, the input differential pair Q_1–Q_2 is loaded in a current mirror formed by Q_3 and Q_4. The second stage is formed by the current-source-loaded common-emitter transistor Q_5. Unlike the CMOS circuit, here there is an output stage formed by the emitter follower Q_6. The function of capacitor C_C will be explained later, in Chapter 11. All transistors have $\beta = 100$, $|V_{BE}| = 0.7$ V, and $r_o = \infty$.

(a) For inputs grounded and output held at 0 V (by negative feedback, not shown) find the emitter currents of all transistors.
(b) Calculate the gain of the amplifier with $R_L = 1$ kΩ.

9.111 A BJT differential amplifier, biased to have $r_e = 100\ \Omega$ and utilizing two 100-Ω emitter resistors and 5-kΩ loads, drives a second differential stage biased to have $r_e = 50\ \Omega$. All BJTs have $\beta = 100$. What is the voltage gain of the first stage? Also find the input resistance of the first stage, and the current gain from the input of the first stage to the collectors of the second stage.

9.112 In the multistage amplifier of Fig. 9.38, emitter resistors are to be introduced: 100 Ω in the emitter lead of each of the first-stage transistors and 25 Ω for each of the second-stage transistors. What is the effect on input resistance, the voltage gain of the first stage, and the overall voltage gain? Use the bias values found in Example 9.7.

D 9.113 Consider the circuit of Fig. 9.38 and its output resistance. Which resistor has the most effect on the output resistance? What should this resistor be changed to if the output resistance is to be reduced by a factor of 2? What will the amplifier gain become after this change? What

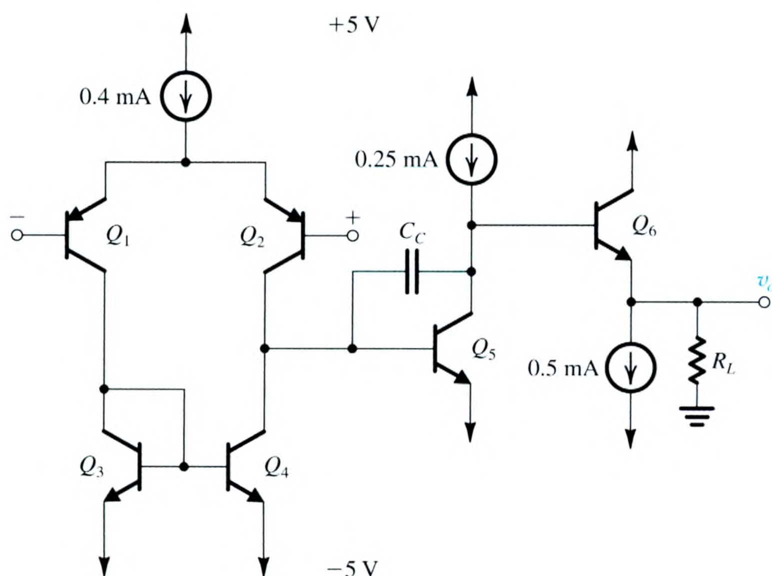


Figure P9.110

other change can you make to restore the amplifier gain to approximately its prior value?

D 9.114 (a) If, in the multistage amplifier of Fig. 9.38, the resistor R_5 is replaced by a constant-current source $\simeq 1$ mA, such that the bias situation is essentially unaffected, what does the overall voltage gain of the amplifier become? Assume that the output resistance of the current source is very high. Use the results of Example 9.8.
(b) With the modification suggested in (a), what is the effect of the change on output resistance? What is the overall gain of the amplifier when loaded by 100 Ω to ground? The original amplifier (before modification) has an output resistance of 152 Ω and a voltage gain of 8513 V/V. What is its gain when loaded by 100 Ω? Comment. Use $\beta = 100$.

***9.115** Figure P9.115 shows a three-stage amplifier in which the stages are directly coupled. The amplifier, however, utilizes bypass capacitors, and, as such, its frequency response falls off at low frequencies. For our purposes here, we shall assume that the capacitors are large enough to act as perfect short circuits at all signal frequencies of interest.

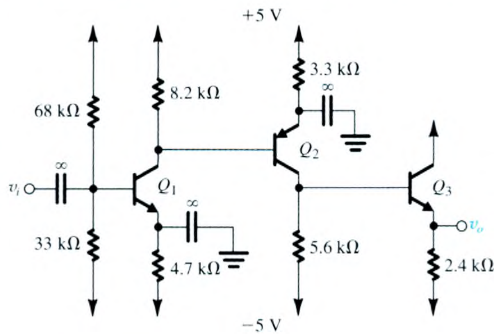


Figure P9.115

(a) Find the dc bias current in each of the three transistors. Also find the dc voltage at the output. Assume $|V_{BE}| = 0.7\,\text{V}$, $\beta = 100$, and neglect the Early effect.
(b) Find the input resistance and the output resistance.
(c) Use the current-gain method to evaluate the voltage gain v_o/v_i.

****9.116** The MOS differential amplifier shown in Fig. P9.116 utilizes three current mirrors for signal transmission: Q_4–Q_6 has a transmission factor of 2 [i.e., $(W/L)_6/(W/L)_4 = 2$], Q_3–Q_5 has a transmission factor of 1, and Q_7–Q_8 has a transmission factor of 2. All transistors are sized to operate at the same overdrive voltage, $|V_{OV}|$. All transistors have the same Early voltage $|V_A|$.

(a) Provide in tabular form the values of I_D, g_m, and r_o of each of the eight transistors in terms of I, V_{OV}, and V_A.
(b) Show that the differential voltage gain A_d is given by

$$A_d = 2g_{m1}(r_{o6} \| r_{o8}) = V_A/V_{OV}$$

(c) Show that the CM gain is given by

$$|A_{cm}| \simeq \frac{r_{o6} \| r_{o8}}{R_{SS}} \frac{1}{g_{m7}r_{o7}}$$

where R_{SS} is the output resistance of the bias current source I. (*Hint*: Since the input differential circuit is balanced, each of Q_1 and Q_2 will conduct a common-mode signal current of approximately $(v_{icm}/2R_{SS})$. These currents are fed into the current mirror loads. Use the current-mirror equivalent circuit derived in Section 8.2.5. Specificially, for each current mirror, the current transfer ratio is given by

$$A_i \simeq A_i\ (\text{ideal}) \left(1 - \frac{1}{g_m r_o}\right)$$

where g_m and r_o are the parameters of the input transistor of the mirror.)

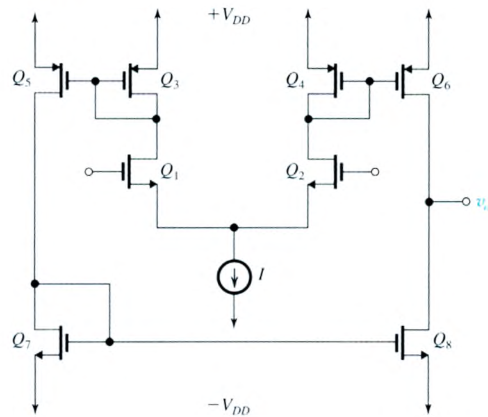


Figure P9.116

(d) If the current source I is implemented using a simple mirror and the MOS transistor is operated at the same V_{OV}, show that the CMRR is given by

$$\text{CMRR} = 4(V_A/V_{OV})^2$$

(e) Find the input CM range and the output linear range in terms of V_{DD}, $|V_t|$, and $|V_{OV}|$.

CHAPTER 10

Frequency Response

IN THIS CHAPTER YOU WILL LEARN

- The internal capacitive effects in the MOSFET and the BJT and how to model these effects by adding capacitances to the hybrid-π model of each transistor type.
- The high-frequency limitation on the gain of CS and CE amplifiers, and how the gain falloff and the upper 3-dB frequency f_H are mostly determined by the small capacitance between the drain and gate (collector and base).
- A powerful method for analyzing the high-frequency response of amplifier circuits of varying complexity.
- How the cascode amplifier studied in Chapter 8 can be designed to obtain wider bandwidth than is possible with CS and CE amplifiers.
- The high-frequency performance of source and emitter followers.
- The high-frequency performance of differential amplifiers.
- Circuit configurations for obtaining wideband amplification.
- How coupling and bypass capacitors cause the gain of discrete-circuit amplifiers to fall off at low frequencies, and how to estimate the frequency f_L at which the gain decreases by 3 dB below its value at midband.

Introduction

In our study of transistor amplifiers in Chapters 7 through 9, we have assumed that their gain is constant independent of the frequency of the input signal. This would imply that their bandwidth is infinite, which of course is not true! To illustrate, we show in Fig. 10.1 a sketch of the magnitude of the gain versus the frequency of the input signal of a discrete-circuit BJT or MOS amplifier. Notice that there is indeed a wide frequency range over which the gain remains almost constant. This is the useful frequency range of operation for the particular amplifier. So far, we have been assuming that our amplifiers are operating in this band, called the middle-frequency band or **midband**. The amplifier is designed so that its midband coincides with the frequency spectrum of the signals it is required to amplify. If this were not the case, the amplifier would distort the frequency spectrum of the input signal, with different components of the input signal being amplified by different amounts.

Figure 10.1 shows that at lower frequencies, the magnitude of the amplifier gain falls off. This occurs because the coupling and bypass capacitors no longer have low impedances. Recall that we assumed that their impedances were small enough to act as short circuits. Although this can be true at midband frequencies, as the frequency of the input signal is lowered, the reactance $1/j\omega C$ of each of these capacitors becomes significant and, as we will show in Section 10.8, this results in a decrease in the overall voltage gain of the amplifier. In the analysis of the low-frequency response of discrete-circuit amplifiers in Section 10.8 we will be particularly interested in how to determine the frequency f_L, which defines the lower end of the midband. It is usually defined as the frequency at which the gain drops by 3 dB below its value in midband. Integrated-circuit amplifiers don't use coupling and bypass capacitors, and so their midband extends down to zero frequency (dc), as shown in Fig. 10.2.

Figures 10.1 and 10.2 indicate that the gain of the amplifier falls off at the high-frequency end. This is due to internal capacitive effects in the transistor. We will study these effects in Section 10.1 and model them with capacitances that we will add to the hybrid-π models. We

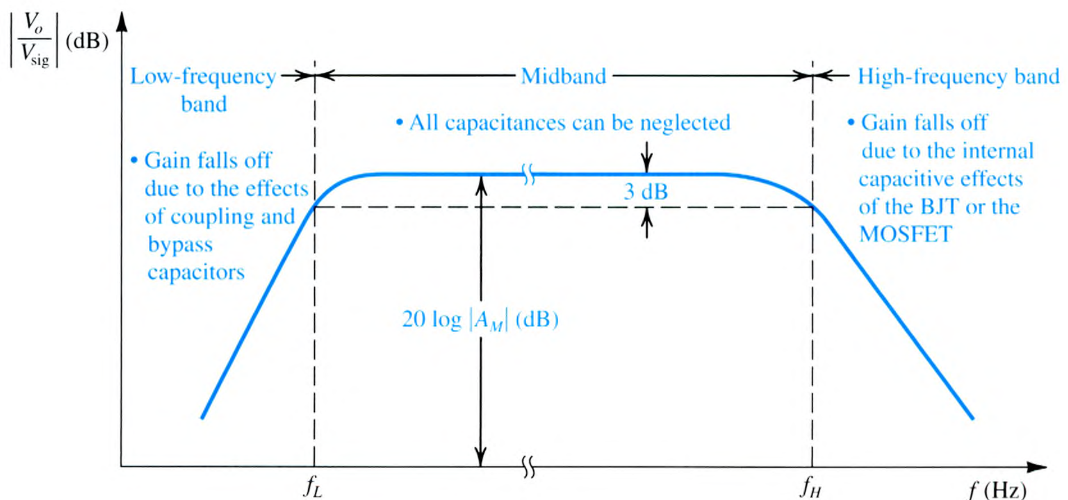


Figure 10.1 Graph of the gain of a discrete-circuit BJT or MOS amplifier versus the frequency of the input signal, showing the three frequency bands relevant to frequency-response determination.

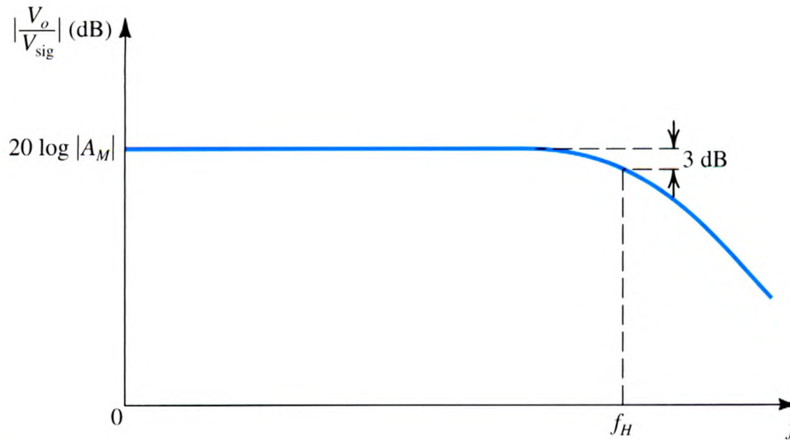


Figure 10.2 Frequency response of a direct-coupled (dc) amplifier. Notice that the gain does *not* fall off at low frequencies, and the midband gain A_M extends down to zero frequency.

will use the resulting high-frequency device models in Section 10.3 in the analysis of the high-frequency response of CS and CE amplifiers. We will be specifically interested in how to determine the frequency f_H, which defines the upper end of the midband. It is defined as the frequency at which the gain drops by 3 dB below its midband value. Since $f_H \gg f_L$, the amplifier bandwidth is usually taken to be equal to f_H,

$$\text{BW} = f_H$$

A figure of merit for the amplifier is its **gain–bandwidth product**, defined as

$$GB = |A_M|\text{BW}$$

We will see that in amplifier design, it is usually possible to trade off gain for bandwidth.

This chapter is concerned with the frequency-response analysis of amplifier configurations of varying degrees of complexity. We are particularly interested in ways to extend the amplifier bandwidth (i.e., increase f_H) either by adding specific circuit components, such as source and emitter-degeneration resistances, or by changing the circuit configuration altogether.

Before going further, we encourage you to review Section 1.6, which introduces amplifier frequency response and the extremely important topic of single-time-constant (STC) circuits. More details on STC circuits can be found in Appendix E. As well, Appendix F provides a review of important tools from circuit and system theory: poles, zeros, and Bode plots.

Finally, a note on notation: Since we will be dealing with quantities that are functions of frequency, or, equivalently, the Laplace variable s, we will be using capital letters with lowercase subscripts for our symbols. This practice conforms with the symbol notation introduced in Chapter 1.

10.1 High-Frequency Transistor Models

The gain falloff at high frequencies is caused by the capacitive effects internal to the transistors. In this section we will briefly consider these effects and, more importantly, show how the transistor small-signal model can be augmented to take these effects into account.

10.1.1 The MOSFET

From our study of the physical operation of the MOSFET in Section 5.1, we know that the device has internal capacitances. In fact, we used one of these, the gate-to-channel capacitance, in our derivation of the MOSFET *i–v* characteristics. There, however, we assumed that the steady-state charges on these capacitances are acquired instantaneously. In other words, we did not account for the finite time required to charge and discharge the various internal capacitances. As a result, the device models we derived, such as the small-signal model, do not include any capacitances. The use of these models would predict constant amplifier gain independent of frequency. We know, however, that this (unfortunately) does not happen in practice; in fact, the gain of every MOSFET amplifier falls off at some high frequency. Similarly, the MOSFET digital logic inverter (Chapter 16) exhibits a finite nonzero propagation delay. To be able to predict these results, we must augment the MOSFET model by including internal capacitances. This is the subject of this section.

To visualize the physical origin of the various internal capacitances, refer to Fig. 10.3, which shows the cross section of an *n*-channel MOSFET operating in the saturation region, as signified by the tapered *n* channel that is pinched off at the drain end. There are four internal capacitances: Two of these, C_{gs} and C_{gd}, result from the gate-capacitance effect; the other two, C_{sb} and C_{db}, are the depletion capacitances of the *pn* junctions formed by the source region and the substrate, and the drain region and the substrate, respectively.

We discussed the gate-capacitive effect in Section 5.1. Briefly, the polysilicon gate forms a parallel-plate capacitor with the channel region, with the oxide layer serving as the capacitor dielectric. The gate (or oxide) capacitance per unit gate area is denoted C_{ox}. When the channel is tapered and pinched off, the gate capacitance is given by $\frac{2}{3}WLC_{ox}$. In addition to this capacitance, there are two other small capacitances resulting from the overlap of the gate with the source region (or source diffusion) and the drain region (or drain diffusion). Each of these overlaps has a length L_{ov}, so the resulting overlap capacitances C_{ov} are given by

$$C_{ov} = WL_{ov}C_{ox} \tag{10.1}$$

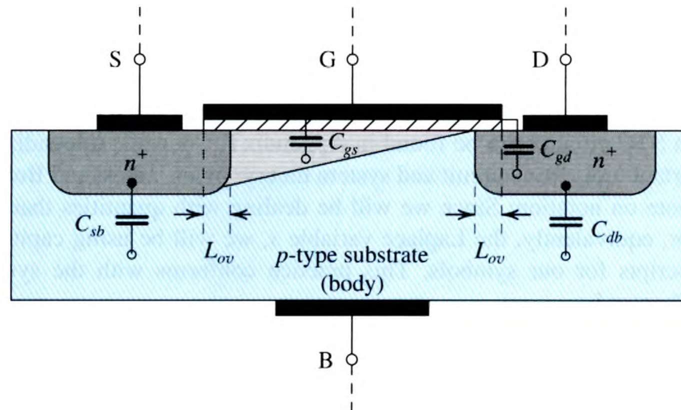


Figure 10.3 Cross section of an *n*-channel MOSFET operating in the saturation region. The four internal capacitances, C_{gs}, C_{gd}, C_{sb}, and C_{db}, are indicated. Note that the bias voltages are not shown. Also not shown, to keep the diagram simple, is the depletion region.

Typically, $L_{ov} = 0.05$ to $0.1\,L$. We can now express the gate-to-source capacitance C_{gs} as

$$C_{gs} = \frac{2}{3} WLC_{ox} + C_{ov} \tag{10.2}$$

For the gate-to-drain capacitance, we note that the channel pinch-off at the drain end causes C_{gd} to consist entirely of the overlap component C_{ov},

$$C_{gd} = C_{ov} \tag{10.3}$$

The depletion-layer capacitances of the two reverse-biased *pn* junctions formed between each of the source and the drain diffusions and the *p*-type substrate (body) can be determined using the formula developed in Section 3.6 (Eq. 3.47). Thus, for the source diffusion, we have the source-body capacitance, C_{sb},

$$C_{sb} = \frac{C_{sb0}}{\sqrt{1 + \dfrac{V_{SB}}{V_0}}} \tag{10.4}$$

where C_{sb0} is the value of C_{sb} at zero body-source bias, V_{SB} is the magnitude of the reverse-bias voltage, and V_0 is the junction built-in voltage (0.6 V to 0.8 V). Similarly, for the drain diffusion, we have the drain-body capacitance C_{db},

$$C_{db} = \frac{C_{db0}}{\sqrt{1 + \dfrac{V_{DB}}{V_0}}} \tag{10.5}$$

where C_{db0} is the capacitance value at zero reverse-bias voltage and V_{DB} is the magnitude of this reverse-bias voltage. Note that we have assumed that for both junctions, the grading coefficient $m = \frac{1}{2}$.

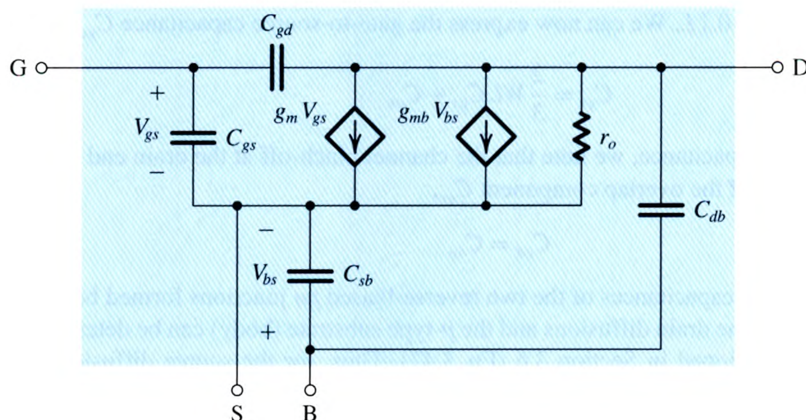
Note that each of these junction capacitances includes a component arising from the bottom side of the diffusion and a component arising from the *sidewalls* of the diffusion. Specifically, each diffusion has three sidewalls that are in contact with the substrate and that thus contribute to the junction capacitance (the fourth wall is in contact with the channel).

The formulas for the junction capacitances in Eqs. (10.4) and (10.5) assume small-signal operation. Typical values for the various capacitances exhibited by an *n*-channel MOSFET in a 0.15-μm CMOS process are given in the following exercise.

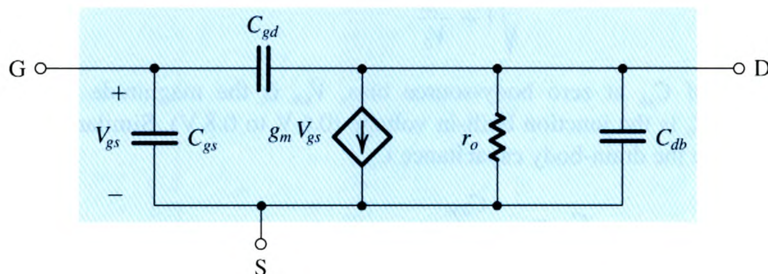
EXERCISE

10.1 For an *n*-channel MOSFET with $t_{ox} = 3$ nm, $L = 0.15\,\mu$m, $W = 1.5\,\mu$m, $L_{ov} = 0.03\,\mu$m, $C_{sb0} = C_{db0} = 1$ fF, $V_0 = 0.9$ V, $V_{SB} = 0.2$ V, and $V_{DS} = 0.8$ V, calculate the following capacitances when the transistor is operating in saturation: C_{ox}, C_{ov}, C_{gs}, C_{gd}, C_{sb}, and C_{db}.

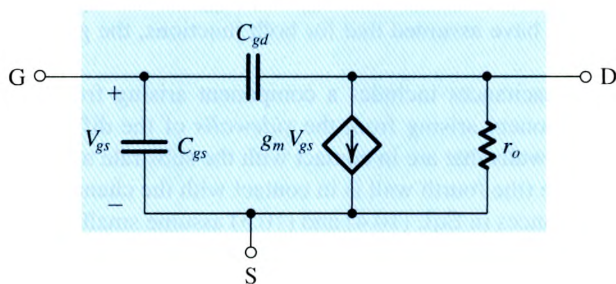
Ans. 11.5 fF/μm^2; 0.52 fF; 2.25 fF; 0.52 fF; 0.9 fF; 0.7 fF



(a)



(b)



(c)

Figure 10.4 (**a**) High-frequency, equivalent-circuit model for the MOSFET. (**b**) The equivalent circuit for the case in which the source is connected to the substrate (body). (**c**) The equivalent-circuit model of (**b**) with C_{db} neglected (to simplify analysis).

The High-Frequency MOSFET Model Figure 10.4(a) shows the small-signal model of the MOSFET, including the four capacitances C_{gs}, C_{gd}, C_{sb}, and C_{db}. This model can be used to predict the high-frequency response of MOSFET amplifiers. Because it is quite complex for manual analysis, its use is limited to computer simulation using SPICE or similar software. Fortunately, when the source is connected to the body, the model simplifies considerably, as shown in Fig. 10.4(b). In this model, C_{gd}, although small, plays a significant role in determining the high-frequency response of amplifiers, so it must be kept in the model.

Capacitance C_{db}, on the other hand, can sometimes be neglected, resulting in significant simplification of manual analysis. The resulting circuit is shown in Fig. 10.4(c).

The MOSFET Unity-Gain Frequency (f_T) A figure of merit for the high-frequency operation of the MOSFET as an amplifier is the unity-gain frequency, f_T, also known as the **transition frequency**, which gives rise to the subscript T. This is defined as *the frequency at which the short-circuit current gain of the common-source configuration becomes unity.* Figure 10.5 shows the MOSFET hybrid-π model with the source as the common terminal between the input and output ports. To determine the short-circuit current gain, the input is fed with a current-source signal I_i and the output terminals are short-circuited. We can see that the current in the short circuit is given by

$$I_o = g_m V_{gs} - sC_{gd}V_{gs} = (g_m - sC_{gd})V_{gs} \tag{10.6}$$

From Fig. 10.5, we see that C_{gd} appears in parallel with C_{gs} and thus we can express V_{gs} in terms of the input current I_i as

$$V_{gs} = \frac{I_i}{s(C_{gs} + C_{gd})} \tag{10.7}$$

Equations (10.6) and (10.7) can be combined to obtain the short-circuit current gain,

$$\frac{I_o}{I_i} = \frac{g_m - sC_{gd}}{s(C_{gs} + C_{gd})} \tag{10.8}$$

For physical frequencies $s = j\omega$, the magnitude of the current gain is

$$\left|\frac{I_o}{I_i}\right| \simeq \frac{g_m}{\omega(C_{gs} + C_{gd})}$$

where we have neglected ωC_{gd} relative to g_m. The gain magnitude becomes one at the frequency

$$\omega_T = g_m/(C_{gs} + C_{gd})$$

Thus the unity-gain frequency $f_T = \omega_T/2\pi$ is

$$f_T = \frac{g_m}{2\pi(C_{gs} + C_{gd})} \tag{10.9}$$

Since f_T is proportional to g_m, which determines the midband gain, and inversely proportional to the MOSFET internal capacitances, which limit the amplifier bandwidth, the higher the value of f_T, the more effective the MOSFET becomes as an amplifier. Substituting

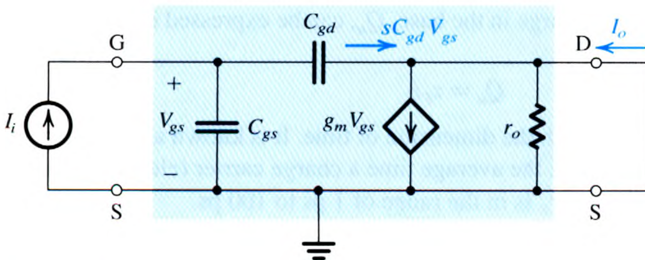


Figure 10.5 Determining the short-circuit current gain I_o/I_i.

for g_m using Eq. (7.41), we can express f_T in terms of the bias current I_D (see Problem 10.3). Alternatively, we can substitute for g_m from Eq. (7.40) to express f_T in terms of the overdrive voltage V_{OV} (see Problem 10.4). Both expressions give us additional insight into the high-frequency operation of the MOSFET. Also refer to Appendix G for a further discussion of f_T.

Typically, f_T ranges from about hundreds of MHz for the older technologies (e.g., a 0.5-μm CMOS process) to hundreds of GHz for newer high-speed technologies (e.g., a 28-nm CMOS process).

Summary For easy reference, the MOSFET high-frequency model, together with expressions for its parameters, is presented in Table 10.1 in the Summary Tables supplement found online at www.oup.com/he/sedra-smith8e.

EXERCISE

10.2 Calculate f_T for the n-channel MOSFET whose capacitances were found in Exercise 10.1. Assume operation at 100 μA and that $k_n' = 500\ \mu A/V^2$.
Ans. 57.5 GHz

10.1.2 The BJT

In our study of the physical operation of the BJT in Section 6.1, we assumed transistor action to be instantaneous, and as a result the transistor models we developed do not include any elements (i.e., capacitors or inductors) that would cause time or frequency dependence. Actual transistors, however, exhibit charge-storage phenomena that limit the speed and frequency of their operation. We have already encountered such effects in our study of the *pn* junction in Chapter 3, where we learned that they can be modeled using capacitances. Here we study the charge-storage effects that take place in the BJT and account for them by adding capacitances to the hybrid-π model. The resulting augmented BJT model will be able to predict the observed dependence of amplifier gain on frequency.

There are three major charge-storage effects in the BJT, each of which can be modeled by a capacitance as follows:

1. **The Base-Charging or Diffusion Capacitance C_{de}.** When the transistor is operating in the active mode, minority-carrier charge is stored in the base region. For an *npn* transistor, the stored electron charge in the base, Q_n, can be expressed in terms of the collector current i_C as

$$Q_n = \tau_F i_C \tag{10.10}$$

where τ_F is a device constant with the dimension of time. It is known as the **forward base-transit time** and represents the average time a charge carrier (electron) spends in crossing the base. Typically, τ_F is in the range of 1 ps to 100 ps.

Equation (10.10) applies for large signals and, since i_C is exponentially related to v_{BE}, Q_n will similarly depend on v_{BE}. For small signals, we can define the

small-signal diffusion capacitance C_{de},

$$C_{de} \equiv \frac{dQ_n}{dv_{BE}} \tag{10.11}$$

$$= \tau_F \frac{di_C}{dv_{BE}}$$

resulting in

$$C_{de} = \tau_F g_m = \tau_F \frac{I_C}{V_T} \tag{10.12}$$

where I_C is the dc collector bias current at which the transistor is operating. Thus, whenever v_{BE} changes by v_{be}, the collector current changes by $g_m v_{be}$ and the charge stored in the base changes by $C_{de} v_{be} = (\tau_F g_m) v_{be}$. This incremental charge has to be supplied by the base current.

2. **The Base–Emitter Junction Capacitance C_{je}.** A change in v_{BE} changes not only the charge stored in the base region but also the charge stored in the base–emitter depletion layer. This distinct charge-storage effect is represented by the EBJ depletion-layer capacitance, C_{je}. From the development in Chapter 3, we know that for a forward-biased junction, which the EBJ is, the depletion-layer capacitance is given approximately by

$$C_{je} \simeq 2C_{je0} \tag{10.13}$$

where C_{je0} is the value of C_{je} at zero EBJ voltage.

3. **The Collector–Base Junction Capacitance C_{μ}.** In active-mode operation, the CBJ is reverse biased, and its junction or **depletion capacitance**, usually denoted C_{μ}, can be found from

$$C_{\mu} = \frac{C_{\mu 0}}{\left(1 + \dfrac{V_{CB}}{V_{0c}}\right)^m} \tag{10.14}$$

where $C_{\mu 0}$ is the value of C_{μ} at zero voltage; V_{CB} is the magnitude of the CBJ reverse-bias voltage, V_{0c} is the CBJ built-in voltage (typically, 0.75 V), and m is its grading coefficient (typically, 0.2–0.5).

The High-Frequency Model Figure 10.6 shows the hybrid-π model of the BJT, including capacitive effects. Specifically, there are two capacitances: the emitter–base

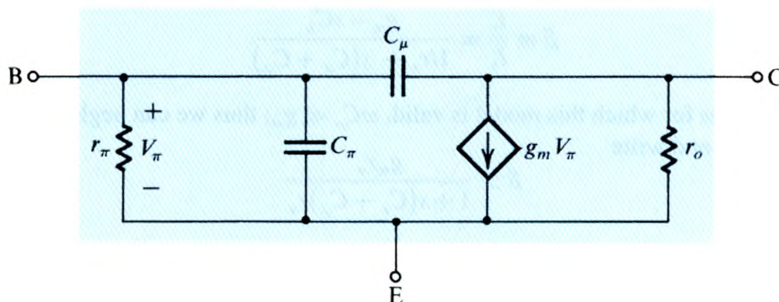


Figure 10.6 The high-frequency hybrid-π model of the BJT.

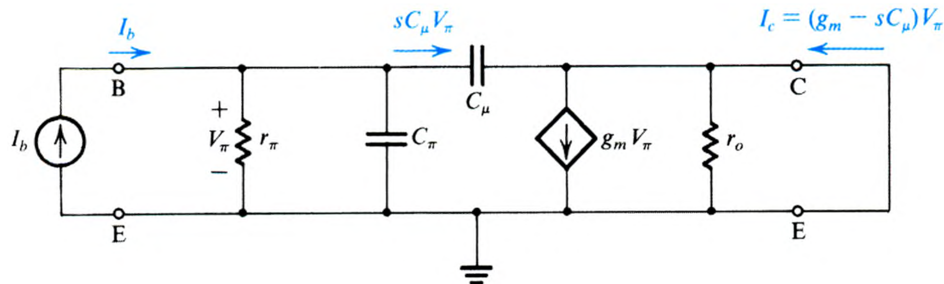


Figure 10.7 Circuit for deriving an expression for $\beta(s) \equiv I_c/I_b$.

capacitance $C_\pi = C_{de} + C_{je}$ and the collector–base capacitance C_μ. Typically, C_π is in the range of a few picofarads to a few tens of picofarads, and C_μ is in the range of a fraction of a picofarad to a few picofarads.[1]

The values of the model parameters can be determined at a given bias point using the formulas presented in this section and in Chapter 6. They can also be found from the terminal measurements specified on the BJT data sheets. For computer simulation, SPICE uses the parameters of the given IC process technology to evaluate the BJT model parameters (see Appendix B).

The BJT Unity-Gain Frequency The transistor data sheets do not usually specify the value of C_π. Rather, they normally give the behavior of β (or h_{fe}) versus frequency. In order to determine C_π and C_μ, we will derive an expression for β, the CE short-circuit current gain, as a function of frequency in terms of the hybrid-π components. For this purpose consider the circuit shown in Fig. 10.7, in which the collector is shorted to the emitter. A node equation at C provides the short-circuit collector current I_c as

$$I_c = (g_m - sC_\mu)V_\pi \tag{10.15}$$

The relationship between V_π and I_b can be established by multiplying I_b by the impedance seen between B and E:

$$V_\pi = I_b(r_\pi \parallel C_\pi \parallel C_\mu) = \frac{I_b}{1/r_\pi + sC_\pi + sC_\mu} \tag{10.16}$$

Thus we can find β by combining Eqs. (10.15) and (10.16):

$$\beta \equiv \frac{I_c}{I_b} = \frac{g_m - sC_\mu}{1/r_\pi + s(C_\pi + C_\mu)} \tag{10.17}$$

At the frequencies for which this model is valid, $\omega C_\mu \ll g_m$; thus we can neglect the sC_μ term in the numerator and write

$$\beta \simeq \frac{g_m r_\pi}{1 + s(C_\pi + C_\mu)r_\pi}$$

[1]These values apply for discrete devices and devices fabricated with a relatively old IC process technology (the so-called high-voltage process; see Appendix K). For modern IC fabrication processes, C_π and C_μ are in the range of femtofarads (fF) to tens of femtofarads.

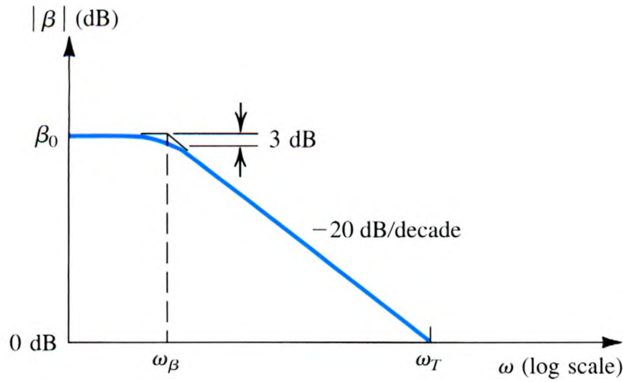


Figure 10.8 Bode plot for $|\beta|$.

Thus,

$$\beta = \frac{\beta_0}{1 + s(C_\pi + C_\mu)r_\pi} \tag{10.18}$$

where β_0 is the low-frequency value of β. Thus β has a single-pole (or single-time-constant) response with a 3-dB frequency at $\omega = \omega_\beta$, where

$$\omega_\beta = \frac{1}{(C_\pi + C_\mu)r_\pi} \tag{10.19}$$

Figure 10.8 shows a Bode plot for $|\beta|$. From the –20-dB/decade slope, it follows that the frequency at which $|\beta|$ drops to unity, which is called the **unity-gain bandwidth** ω_T, is given by

$$\omega_T = \beta_0 \omega_\beta$$

Thus,

$$\omega_T = \frac{g_m}{C_\pi + C_\mu} \tag{10.20}$$

and

$$f_T = \frac{g_m}{2\pi(C_\pi + C_\mu)} \tag{10.21}$$

This expression is identical to that of f_T for the MOSFET (Eq. 10.9) with C_π replacing C_{gs} and C_μ replacing C_{gd}.

The unity-gain bandwidth f_T, also known as the **transition frequency**, which gives rise to the subscript T, is usually specified on the data sheets of a transistor. In some cases f_T is given as a function of I_C and V_{CE}. To see how f_T changes with I_C, recall that g_m is directly proportional to I_C, but only part of C_π (the diffusion capacitance C_{de}) is directly proportional to I_C. It follows that f_T decreases at low currents, as shown in Fig. 10.9. However, the decrease in f_T at high currents, also shown in Fig. 10.9, cannot be explained by this argument; rather, it is due to the same phenomenon that causes β_0 to decrease at high currents (Section 6.4.2). In the region where f_T is almost constant, C_π is dominated by the diffusion part and is

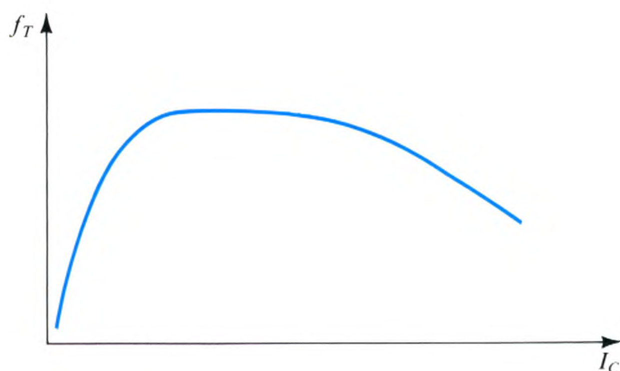


Figure 10.9 Variation of f_T with I_C.

much greater than C_μ. That is, $C_\pi + C_\mu \simeq C_{de} = \tau_F g_m$ and

$$f_T \simeq \frac{1}{2\pi \tau_F} \tag{10.22}$$

Typically, f_T is in the range of hundreds of megahertz to hundreds of gigahertz. The value of f_T can be used in Eq. (10.21) to determine $C_\pi + C_\mu$. The capacitance C_μ is usually determined separately by measuring the capacitance between base and collector at the desired reverse-bias voltage V_{CB}.

Before leaving this section, we should mention that the high-frequency model of Fig. 10.7 characterizes transistor operation fairly accurately up to a frequency of about 0.2 f_T. At higher frequencies one has to add other parasitic elements to the model. In particular, there is another somewhat important capacitive effect in BJTs on an IC chip: the collector-substrate capacitance C_{cb}. We will take C_{cb} into account in some applications.

Summary For easy reference, the BJT high-frequency model, together with formulas for determining its parameter values, is presented in Table 10.2 in the Summary Tables supplement found online at www.oup.com/he/sedra-smith8e.

EXERCISES

10.3 Find C_{de}, C_{je}, C_π, C_μ, and f_T for a BJT operating at a dc collector current $I_C = 1$ mA and a CBJ reverse bias of 2 V. The device has $\tau_F = 20$ ps, $C_{je0} = 20$ fF, $C_{\mu 0} = 20$ fF, $V_{0e} = 0.9$ V, $V_{0c} = 0.5$ V, and $m_{CBJ} = 0.33$.
Ans. 0.8 pF; 40 fF; 0.84 pF; 12 fF; 7.47 GHz

10.4 For a BJT operated at $I_C = 1$ mA, find f_T and C_π if $C_\mu = 2$ pF and $|\beta| = 10$ at 50 MHz.
Ans. 500 MHz; 10.7 pF

10.5 If C_π of the BJT in Exercise 10.4 includes a relatively constant depletion-layer capacitance of 2 pF, find f_T of the BJT when operated at $I_C = 0.1$ mA.
Ans. 130.7 MHz

10.2 High-Frequency Response of CS and CE Amplifiers

Equipped with equivalent-circuit models that represent the high-frequency operation of the MOSFET and the BJT, we now address the question of the high-frequency performance of the CS and CE amplifiers. Our objective is to identify the mechanism that limits the high-frequency performance of these important amplifier configurations. As well, we need to find a simple approach to estimate the frequency f_H at which the gain falls by 3 dB below its value at midband frequencies, $|A_M|$.

The analysis presented here applies equally well to discrete-circuit, capacitively coupled amplifiers and to IC amplifiers, whose frequency responses were shown in Fig. 10.1 and in Fig. 10.2, respectively. At the frequencies of interest to us here (the high-frequency band), all coupling and bypass capacitors behave as perfect short circuits, and amplifiers of both types have identical high-frequency equivalent circuits.

10.2.1 Frequency Response of the Low-Pass Single-Time-Constant Circuit

Analyzing the high-frequency response of amplifiers is much easier once we know the frequency response of the low-pass single-time-constant circuit, shown in its Thévenin form in Fig. 10.10(a) and in its Norton form in Fig. 10.10(b). The circuit has the time constant

$$\tau = CR \qquad (10.23)$$

and the transfer function

$$\frac{V_o}{V_i} = \frac{1}{1+\dfrac{s}{\omega_P}} \qquad (10.24)$$

where the pole-frequency ω_P is given by

$$\omega_P = \frac{1}{\tau} \qquad (10.25)$$

For physical frequencies, $s = j\omega$ and the magnitude of the transfer function is

$$\left|\frac{V_o}{V_i}\right| = \frac{1}{\sqrt{1+\left(\dfrac{\omega}{\omega_P}\right)^2}} \qquad (10.26)$$

Thus, at $\omega = \omega_P$, $|V_o/V_i| = 1/\sqrt{2}$; in other words, the gain drops 3 dB below its value at dc. It follows that the 3-dB frequency ω_H is equal to ω_P.

$$\omega_H = \omega_P \qquad (10.27)$$

For $\omega \gg \omega_P$,

$$\left|\frac{V_o}{V_i}\right| \simeq \frac{\omega_P}{\omega} \qquad (10.28)$$

Thus, whenever ω increases by a factor of 10 (a decade increase), the gain decreases by a factor of 10 (i.e., by 20 dB). This is confirmed by the Bode plot for $|V_o/V_i|$ shown in Fig. 10.10(c).

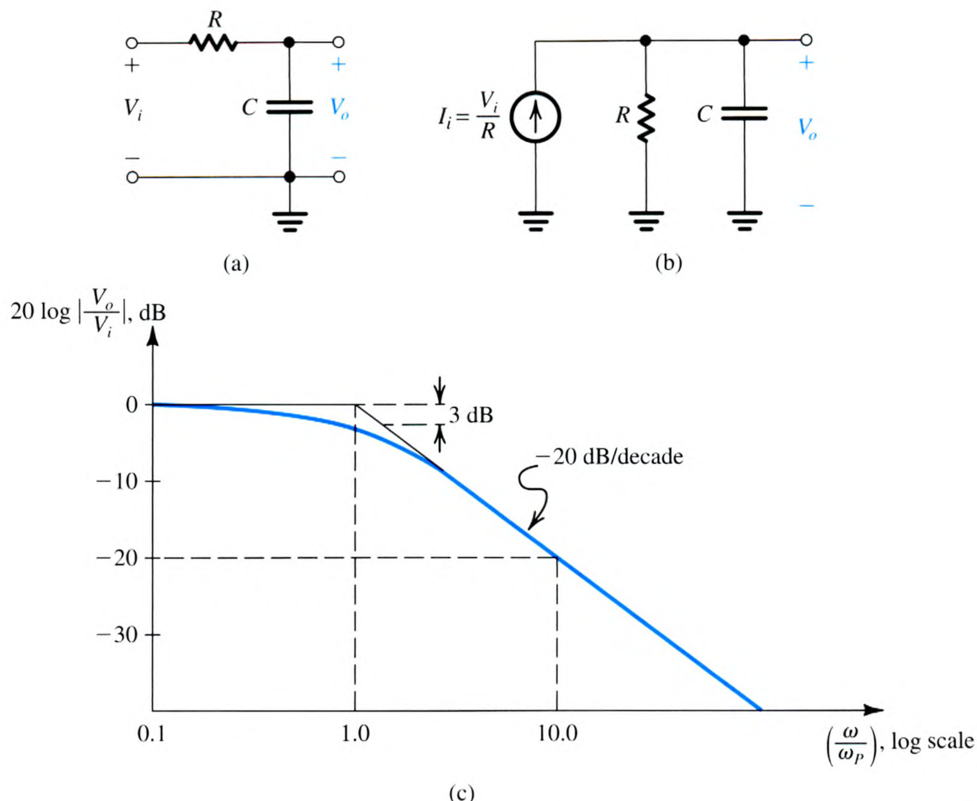


Figure 10.10 The low-pass single-time-constant circuit in its (a) Thévenin form and (b) Norton form. (c) Bode plot for $|V_o/V_i|$.

10.2.2 The Common-Source Amplifier

Figure 10.11(a) shows a common-source amplifier fed with a signal source V_{sig} having a resistance R_{sig} and connected to a load resistance R_L. Here R_L represents either a passive or an active (i.e., current source) load. Replacing the MOSFET with its high-frequency hybrid-π equivalent-circuit model results in the circuit shown in Fig. 10.11(b). This circuit can be simplified as illustrated in Fig. 10.11(c), by combining R_L and r_o in a single resistance R_L',

$$R_L' = R_L \| r_o \tag{10.29}$$

The midband gain A_M can be found from the circuit in Fig. 10.11(c) by setting $C_{gs} = C_{gd} = 0$. The result is

$$A_M = \frac{V_o}{V_{sig}} = -g_m R_L' \tag{10.30}$$

We can analyze the circuit in Fig. 10.11(c) to find its transfer function V_o/V_{sig} in terms of the complex frequency variable s. Because two capacitors, C_{gs} and C_{gd}, are present, the resulting transfer function will be of second order. The poles and zeros can then be found. This, however, will not provide us with simple expressions that reveal the essence of what

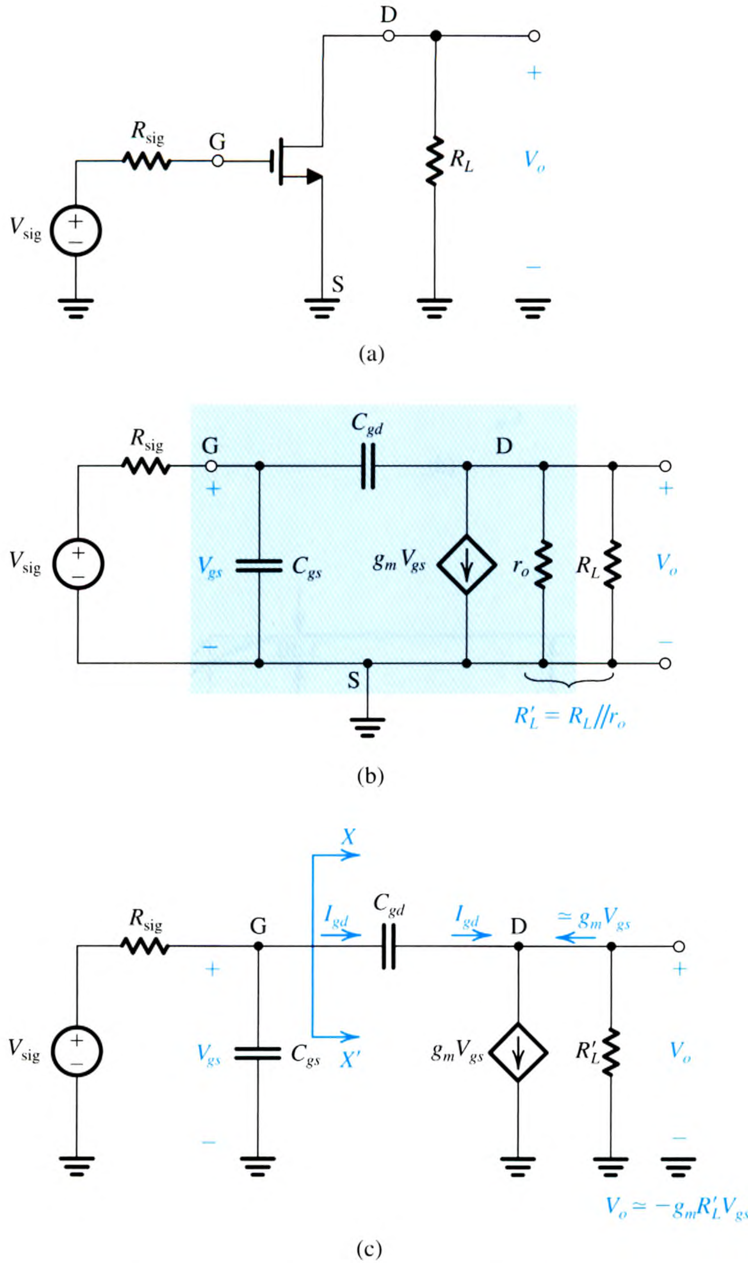
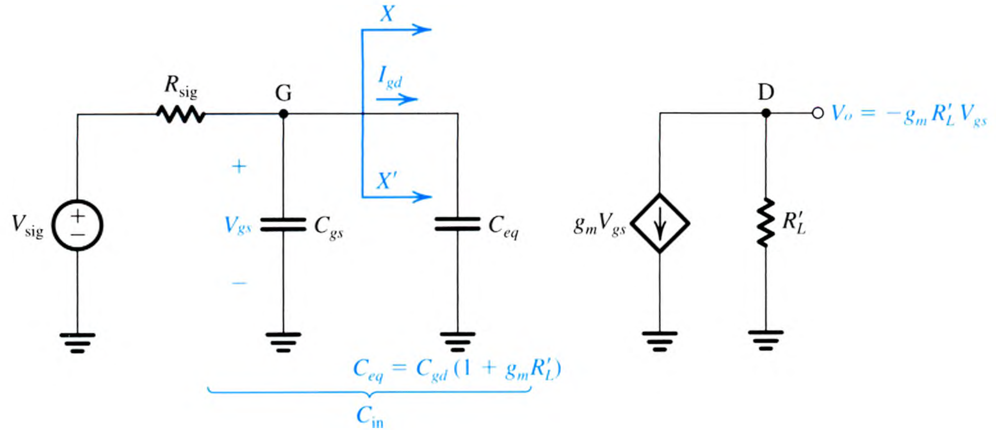
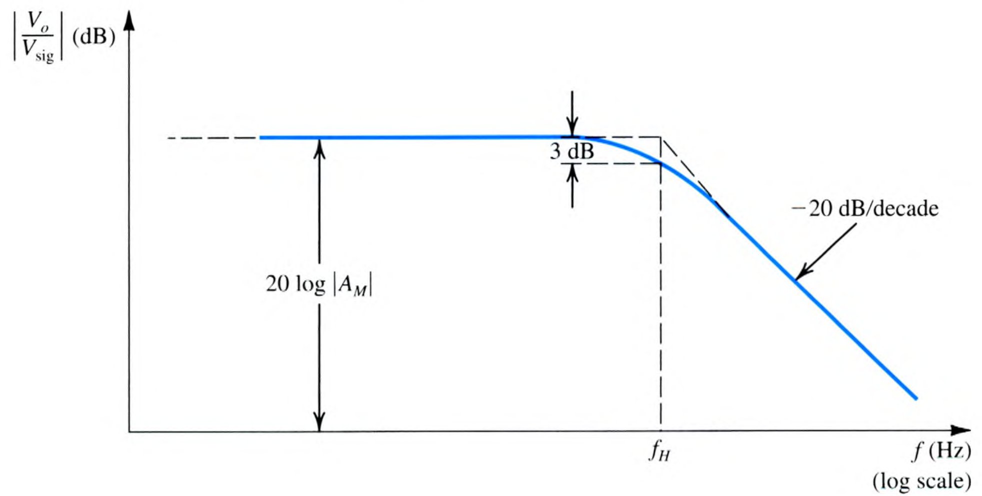


Figure 10.11 Determining the high-frequency response of the CS amplifier in (**a**); (**b**) equivalent circuit; (**c**) the circuit of (**b**) simplified by combining R_L and r_o into a single resistance R'_L; (**d**) the equivalent circuit with C_{gd} replaced at the input side with the equivalent capacitance C_{eq}; (**e**) the frequency-response plot, which is that of a low-pass, single-time-constant circuit.

limits the high-frequency operation of the CS amplifier. We need that insight in order to be able to make intelligent decisions when designing the circuit. Therefore, we will not derive the transfer function and instead we will opt for an approximate approach that will reveal considerable information about the high-frequency operation of the CS amplifier.



(d)



(e)

Figure 10.11 *continued*

Our approach is to focus on the input side of the circuit and simplify the input circuit to a simple RC low-pass network. To do so, we need to replace the bridging capacitor C_{gd} by an equivalent capacitance C_{eq} between node G and ground. To that end, consider first the output node. We can see that the load current is $(g_m V_{gs} - I_{gd})$, where $(g_m V_{gs})$ is the output current of the transistor and I_{gd} is the current supplied through the very small capacitance C_{gd}. At frequencies in the vicinity of f_H, which defines the edge of the midband, we can assume that I_{gd} is still much smaller than $(g_m V_{gs})$, so that V_o can be given approximately by

$$V_o \simeq -(g_m V_{gs})R'_L = -g_m R'_L V_{gs} \tag{10.31}$$

Eq. (10.31) indicates that the gain from gate to drain is $-g_m R_L'$, the same value as in the midband. The current I_{gd} can now be found as

$$\begin{aligned} I_{gd} &= sC_{gd}(V_{gs} - V_o) \\ &= sC_{gd}[V_{gs} - (-g_m R_L' V_{gs})] \\ &= sC_{gd}(1 + g_m R_L')V_{gs} \end{aligned}$$

Now, the left-hand side of the circuit in Fig. 10.11(c), at XX', knows of the existence of C_{gd} only through the current I_{gd}. Therefore, we can replace C_{gd} by an equivalent capacitance C_{eq} between the gate and ground as long as C_{eq} draws a current equal to I_{gd}. That is,

$$sC_{eq}V_{gs} = sC_{gd}(1 + g_m R_L')V_{gs}$$

which gives us

$$C_{eq} = C_{gd}(1 + g_m R_L') \tag{10.32}$$

Thus C_{gd} gives rise to a much larger capacitance C_{eq}, which appears at the amplifier input. The multiplication effect that C_{gd} undergoes comes about because it is connected between circuit nodes G and D, whose voltages are related by a large negative gain ($-g_m R_L'$). This effect is known as the **Miller effect**, and $(1 + g_m R_L')$ is known as the **Miller multiplier**.

Using C_{eq} enables us to simplify the equivalent circuit at the input side to that shown in Fig. 10.11(d), which we recognize as a single-time-constant (STC) circuit of the low-pass type. Thus we can express the output voltage V_{gs} of the STC circuit as

$$V_{gs} = \frac{1}{1 + \dfrac{s}{\omega_P}} V_{sig} \tag{10.33}$$

where ω_P is the pole frequency of the STC circuit,

$$\omega_P = 1/C_{in}R_{sig} \tag{10.34}$$

with

$$C_{in} = C_{gs} + C_{eq} = C_{gs} + C_{gd}(1 + g_m R_L') \tag{10.35}$$

Combining Eqs. (10.31) and (10.33) results in the following expression for the high-frequency gain of the CS amplifier,

$$\frac{V_o}{V_{sig}} = -\frac{g_m R_L'}{1 + \dfrac{s}{\omega_P}} \tag{10.36}$$

which is in the form

$$\frac{V_o}{V_{sig}} = \frac{A_M}{1 + \dfrac{s}{\omega_H}} \tag{10.37}$$

where the midband gain A_M is given by Eq. (10.30) and ω_H is the upper 3-dB frequency,

$$\omega_H = \omega_P = \frac{1}{C_{in}R_{sig}} \tag{10.38}$$

and

$$f_H = \frac{\omega_H}{2\pi} = \frac{1}{2\pi C_{in}R_{sig}} \tag{10.39}$$

We thus see that the high-frequency response is that of a low-pass STC network with a 3-dB frequency f_H determined by the time constant $C_{in}R_{sig}$. Figure 10.11(e) shows a sketch of the magnitude of the high-frequency gain.

Before leaving this section we want to highlight the following points:

1. The upper 3-dB frequency is determined by the interaction of R_{sig} and $C_{in} = C_{gs} + C_{gd}(1 + g_m R_L')$. It follows that a large value of R_{sig} will cause f_H to be lowered.
2. The total input capacitance C_{in} is usually dominated by C_{eq}, which in turn is made large by the multiplication effect that C_{gd} undergoes. So, although C_{gd} is usually a very small capacitance, its effect on the amplifier frequency response can be very significant as a result of its multiplication by the factor $(1 + g_m R_L')$, which is approximately equal to the midband gain of the amplifier. This is the Miller effect, which causes the CS amplifier to have a large total input capacitance C_{in} and hence a low f_H.
3. To extend the high-frequency response of a MOSFET amplifier, we have to find configurations in which the Miller effect is absent or at least reduced. We will return to this subject at great length in Section 10.4 and beyond.
4. The above analysis, resulting in an STC or a single-pole response, is approximate. Specifically, it is based on neglecting I_{gd} relative to $g_m V_{gs}$, an assumption that applies well at frequencies not too much higher than f_H. An exact analysis of the circuit in Fig. 10.11(c) reveals that the circuit has a second pole with a frequency much greater than f_H, and transmission zeros at $s = \infty$ and $s = g_m/C_{gd}$; the latter's frequency is also much greater than f_H. Thus both the second pole and the zero will have negligible effect on our estimate of f_H. So, the method that uses the Miller effect is more than sufficient for a quick estimate of f_H. As well, the approximate approach helps to reveal the primary limitation on the high-frequency response: the Miller effect.
5. The CS amplifier is said to have a *dominant* high-frequency pole with frequency $f_P \simeq f_H$.

Example 10.1

Find the midband gain A_M and the upper 3-dB frequency f_H of an integrated-circuit CS amplifier fed with a signal source having an internal resistance $R_{sig} = 20\ \text{k}\Omega$. The amplifier has $R_L = 20\ \text{k}\Omega$, $g_m = 2$ mA/V, $r_o = 20\ \text{k}\Omega$, $C_{gs} = 20$ fF, and $C_{gd} = 5$ fF. Also, find the frequency of the transmission zero.

Solution

$$A_M = -g_m R_L'$$

where

$$R_L' = R_L \| r_o = 20 \| 20 = 10\ \text{k}\Omega$$

$$g_m R_L' = 2 \times 10 = 20\ \text{V/V}$$

Thus,

$$A_M = -20\ \text{V/V}$$

The equivalent capacitance, C_{eq}, is found as

$$C_{eq} = (1 + g_m R'_L)C_{gd}$$
$$= (1 + 20) \times 5 = 105 \text{ fF}$$

The total input capacitance C_{in} can be now obtained as

$$C_{in} = C_{gs} + C_{eq} = 20 + 105 = 125 \text{ fF}$$

The upper 3-dB frequency f_H is found from

$$f_H = \frac{1}{2\pi C_{in} R_{sig}}$$
$$= \frac{1}{2\pi \times 125 \times 10^{-15} \times 20 \times 10^3}$$
$$= 63.7 \text{ MHz}$$

Finally, the transmission zero has a frequency

$$f_Z = \frac{g_m}{2\pi C_{gd}} = \frac{2 \times 10^{-3}}{2\pi \times 5 \times 10^{-15}} = 63.7 \text{ GHz}$$

which is 1000 times higher than f_H.

Video Example VE 10.1 Gain and High-Frequency Response of a CS Amplifier

A common-source amplifier is fed with a signal source having a resistance $R_{sig} = 100$ kΩ and feeds a load resistance $R_L = 25$ kΩ. The MOSFET is operating at $g_m = 5$ mA/V and has $r_o = 25$ kΩ. The device capacitances are $C_{gs} = 30$ fF and $C_{gd} = 10$ fF. Find the midband gain A_M, the 3-dB frequency f_H, and the frequency of the transmission zero.

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.


VE 10.1

Related end-of-chapter problem: 10.15

EXERCISES

10.6 For the CS amplifier specified in Example 10.1, find the values of A_M and f_H that result when the load resistance is reduced to 10 kΩ.
Ans. –13.3 V/V; 86.9 MHz

10.7 If it is possible to replace the MOSFET used in the amplifier in Example 10.1 with another having the same C_{gs} but a smaller C_{gd}, what is the maximum value that its C_{gd} can be in order to obtain an f_H of at least 100 MHz?
Ans. 2.8 fF

10.2.3 Frequency Response of the CS Amplifier When R_{sig} Is Low

There are applications in which the CS amplifier is fed with a low-resistance signal source. In such cases, the high-frequency gain will no longer be limited by the interaction of the source resistance and the input capacitance. Rather, the high-frequency limitation occurs at the amplifier output as we will now show.

Figure 10.12(a) shows the high-frequency equivalent circuit of the common-source amplifier in the limiting case when R_{sig} is zero. We have included a capacitance C_L across the load R_L in anticipation that a capacitance at the output node, even if it is small, will play an important role in this case. Also, there always is some capacitance between the output node and ground. This can include C_{db} of the MOSFET, the input capacitance of another amplifier stage our amplifier feeds, other stray capacitances, and so on. Finally, note that we did not include C_L in the previous analysis because its role is not significant when R_{sig} is large.

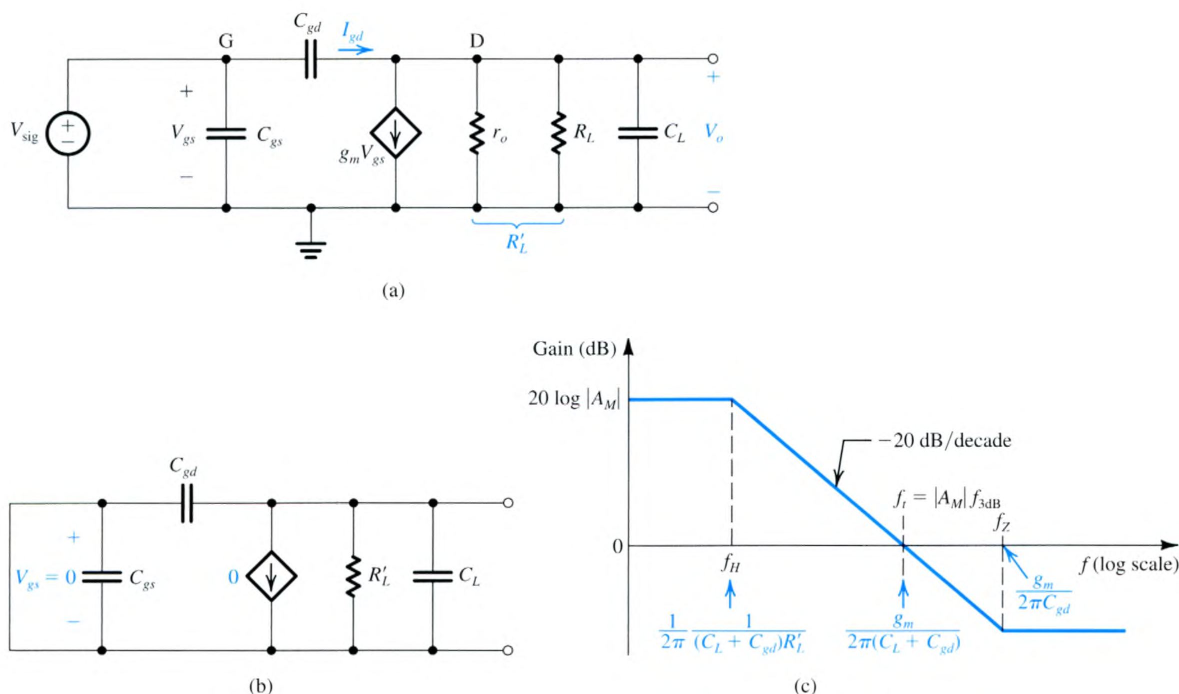


Figure 10.12 **(a)** High-frequency equivalent circuit of a CS amplifier fed with a signal source having a very low (effectively zero) resistance. **(b)** The circuit with V_{sig} reduced to zero to determine the pole. **(c)** Bode plot for the gain of the circuit in **(a)**.

Returning to the circuit in Fig. 10.12(a), we can now derive its transfer function V_o/V_{sig}. First, note that

$$V_{gs} = V_{sig} \tag{10.40}$$

Second, we can express the current I_{gd} that flows through C_{gd} as

$$I_{gd} = sC_{gd}(V_{gs} - V_o) \tag{10.41}$$

Next, we can write a node equation at the output node as

$$I_{gd} = g_m V_{gs} + \frac{V_o}{R_L'} + sC_L V_o \tag{10.42}$$

where

$$R_L' = R_L \| r_o$$

Combining Eqs. (10.41) and (10.42) to eliminate I_{gd}, substituting $V_{gs} = V_{sig}$ from Eq. (10.40), and manipulating the resulting equation to obtain the transfer function V_o/V_{sig} gives us

$$\frac{V_o}{V_{sig}} = -g_m R_L' \frac{1 - s(C_{gd}/g_m)}{1 + s(C_L + C_{gd})R_L'} \tag{10.43}$$

Thus, while the dc gain remains equal to $g_m R_L'$, and the frequency of the transmission zero remains unchanged at

$$f_Z = \frac{g_m}{2\pi\, C_{gd}} \tag{10.44}$$

the high-frequency response is now determined by a pole formed by $(C_L + C_{gd})$ together with R_L'. Thus the 3-dB frequency f_H is now given by

$$f_H = \frac{1}{2\pi\,(C_L + C_{gd})R_L'} \tag{10.45}$$

To see how this pole is formed, refer to Fig. 10.12(b), which shows the equivalent circuit with V_{sig} reduced to zero. Observe that the circuit reduces to a capacitance $(C_L + C_{gd})$ in parallel with a resistance R_L'.

We note that the transmission zero frequency, given by Eq. (10.44), is much higher than f_H,

$$\frac{f_Z}{f_H} = (g_m R_L')\left(1 + \frac{C_L}{C_{gd}}\right) \tag{10.46}$$

Thus, f_Z does not play a significant role in the vicinity of f_H. In fact, the gain decreases from its low-frequency value of $(g_m R_L')$ at a uniform rate of –20 dB/decade, reaching unity (0 dB) at a frequency f_t, which is equal to the **gain–bandwidth product**,

$$\begin{aligned} f_t &= |A_M|\, f_H \\ &= g_m R_L' \frac{1}{2\pi\,(C_L + C_{gd})R_L'} \end{aligned}$$

Thus,

$$f_t = \frac{g_m}{2\pi\,(C_L + C_{gd})} \tag{10.47}$$

Figure 10.12(c) shows a sketch of the high-frequency gain of the CS amplifier.

Example 10.2

Consider an IC CS amplifier fed with a source having $R_{sig} = 0$ and having an effective load resistance R'_L composed of r_o of the amplifier transistor in parallel with an equal resistance r_o of the current-source load. Let $g_m = 2$ mA/V, $r_o = 20\text{k}\Omega$, $C_{gs} = 20$ fF, $C_{gd} = 5$ fF, and $C_L = 25$ fF. Find A_M, f_H, f_t, and f_Z. If the amplifying transistor is to be operated at twice the original overdrive voltage while W and L remain unchanged, by what factor must the bias current be changed? What are the new values of A_M, f_H, f_t, and f_Z?

Solution

The low-frequency gain A_M is given by

$$A_M = -g_m R'_L = -g_m (r_o \| r_o)$$

Thus,

$$A_M = -\frac{1}{2} g_m r_o = -\frac{1}{2} \times 2 \times 20$$
$$= -20 \text{ V/V}$$

The 3-dB frequency f_H can be found using Eq. (10.45),

$$f_H = \frac{1}{2\pi (C_L + C_{gd}) R'_L}$$
$$= \frac{1}{2\pi (25 + 5) \times 10^{-15} \times 10 \times 10^3}$$
$$= 530.5 \text{ MHz}$$

and the unity-gain frequency, which in this case is equal to the gain–bandwidth product, can be determined as

$$f_t = |A_M| f_H = 20 \times 530.5 = 10.6 \text{ GHz}$$

The frequency of the zero is obtained using Eq. (10.44) as

$$f_Z = \frac{1}{2\pi} \frac{g_m}{C_{gd}}$$
$$= \frac{1}{2\pi} \frac{2 \times 10^{-3}}{5 \times 10^{-15}} \simeq 63.7 \text{ GHz}$$

Now, to double V_{OV}, I_D must be quadrupled. The new values of g_m and R'_L can be found as follows:

$$g_m = \frac{I_D}{V_{OV}/2} = 4 \text{ mA/V}$$

$$R'_L = \frac{1}{4} \times 10 = 2.5 \text{ k}\Omega$$

Thus the new value of A_M becomes

$$A_M = -g_m R_L' = -4 \times 2.5 = -10 \text{ V/V}$$

and f_H becomes

$$f_H = \frac{1}{2\pi (C_L + C_{gd}) R_L'}$$

$$= \frac{1}{2\pi (25+5) \times 10^{-15} \times 2.5 \times 10^3}$$

$$= 2.12 \text{ GHz}$$

and the unity-gain frequency (i.e., the gain–bandwidth product) becomes

$$f_t = 10 \times 2.12 = 21.2 \text{ GHz}$$

We note that doubling V_{OV} reduces the dc gain by a factor of 2 and increases the bandwidth by a factor of 4. Thus, the gain–bandwidth product is doubled—a good bargain!

Finally, the frequency of the transmission zero f_Z will be doubled, becoming 127.4 GHz.

EXERCISE

10.8 For the CS amplifier considered in Example 10.2 operating at the original values of V_{OV} and I_D, find the value to which C_L should be increased to place f_t at 3 GHz.
Ans. 101 fF

We conclude this section by noting that the most general case when R_{sig} is not zero and C_L is present will be dealt with in the Section 10.3.

10.2.4 The Common-Emitter Amplifier

Fig. 10.13(a) shows a CE amplifier; its high-frequency equivalent circuit is given in Fig. 10.13(b). The latter circuit can be simplified by using Thévenin's theorem at the input side and combining the two parallel resistances at the output side. Thus, we obtain the equivalent circuit in Fig. 10.13(c) from which we can find the midband gain A_M by setting C_π and C_μ to zero. The result is

$$A_M = \frac{V_o}{V_{sig}} = -\left(\frac{r_\pi}{r_\pi + R_{sig}}\right)(g_m R_L') \tag{10.48}$$

where

$$R_L' = R_L \parallel r_o \tag{10.49}$$

Next we observe that the circuit in Fig. 10.13(c) is identical to that of the CS amplifier in Fig. 10.11(c). Thus the analysis can follow the same process we used for the CS case,

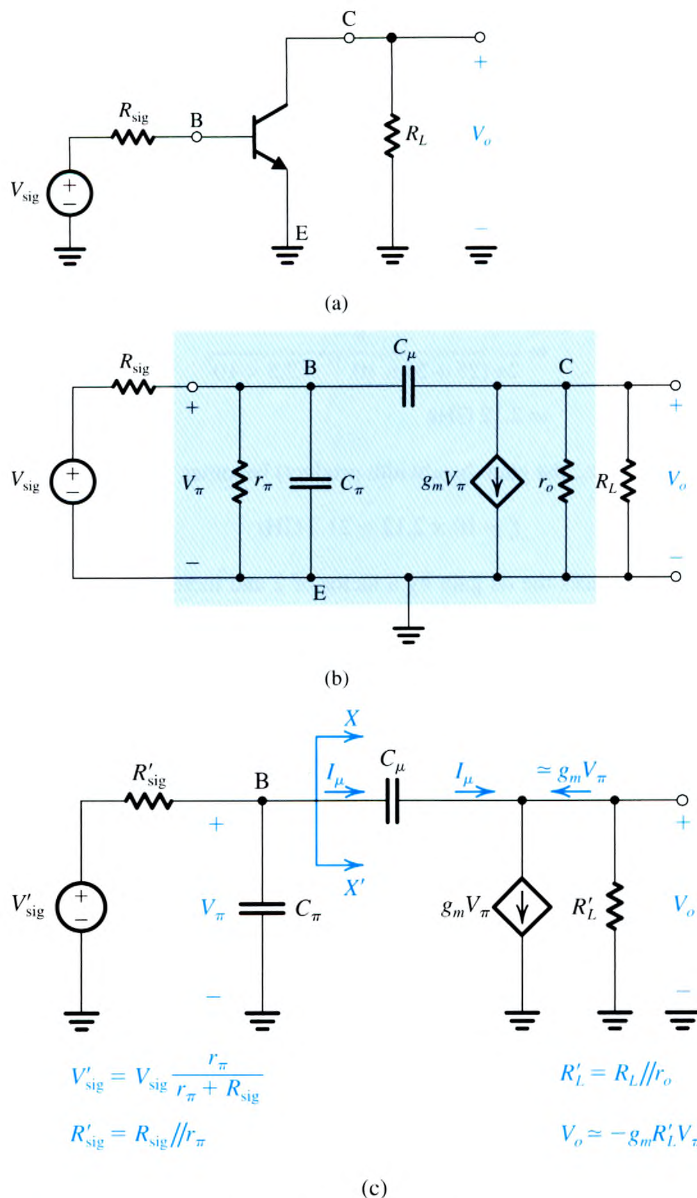
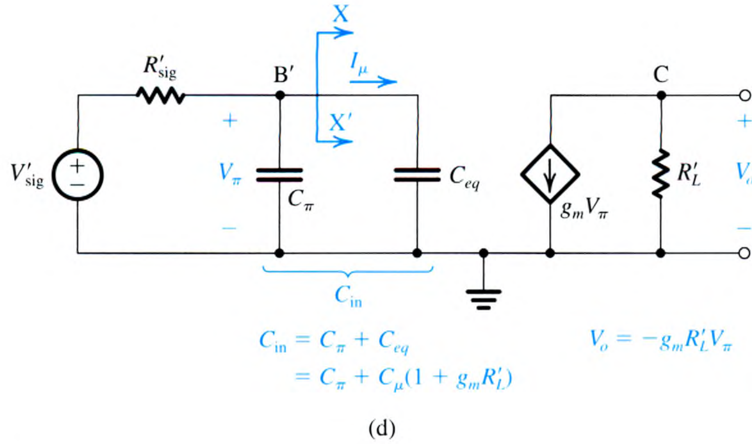


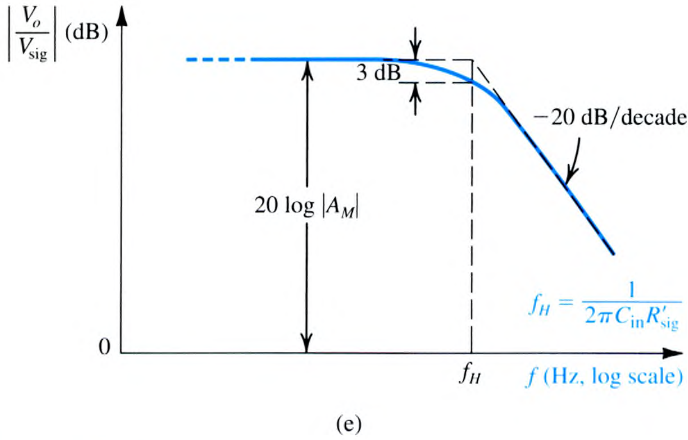
Figure 10.13 Determining the high-frequency response of the CE amplifier: (**a**) amplifier circuit; (**b**) equivalent circuit; (**c**) the circuit of (**b**) simplified at both the input side and the output side; (**d**) equivalent circuit with C_μ replaced at the input side with the equivalent capacitance C_{eq}; (**e**) sketch of the frequency-response Bode plot, which is that of a low-pass STC circuit.

as illustrated in Figs. 10.13(c) and (d). The result is that the CE amplifier gain at high frequencies is given approximately by

$$\frac{V_o}{V_{sig}} = \frac{A_M}{1 + \dfrac{s}{\omega_H}} \tag{10.50}$$



(d)



(e)

Figure 10.13 *continued*

where A_M is given by Eq. (10.48) and the 3-dB frequency f_H is given by

$$f_H = \frac{\omega_H}{2\pi} = \frac{1}{2\pi C_{in} R'_{sig}} \tag{10.51}$$

where

$$C_{in} = C_\pi + C_\mu(1 + g_m R'_L) \tag{10.52}$$

and

$$R'_{sig} = R_{sig} \parallel r_\pi \tag{10.53}$$

Notice that C_{in} is simply the sum of C_π and the Miller capacitance $C_\mu(1 + g_m R'_L)$. From the circuit in Fig. 10.13(b) we can easily find the resistance R'_{sig} seen by C_{in} by reducing V_{sig} to zero, "grabbing hold" of the terminals B and E, and looking back (to the left). There we will see r_π in parallel with R_{sig}.

Finally, comments very similar to those made on the high-frequency response of the CS amplifier can be made here as well.

Example 10.3

Find the midband gain and the upper 3-dB frequency of the common-emitter amplifier of Fig. 10.13(a) for the following case: $I_E = 1$ mA, $R_{sig} = 5$ kΩ, $R_L = 5$ kΩ, $\beta_0 = 100$, $V_A = 100$ V, $C_\mu = 1$ pF, and $f_T = 800$ MHz. Also, determine the frequency of the transmission zero.

Solution

The transistor is biased at $I_C \simeq 1$ mA. Thus the values of its hybrid-π model parameters are

$$g_m = \frac{I_C}{V_T} = \frac{1 \text{ mA}}{0.025 \text{ V}} = 40 \text{ mA/V}$$

$$r_\pi = \frac{\beta_0}{g_m} = \frac{100}{40 \text{ mA/V}} = 2.5 \text{ k}\Omega$$

$$r_o = \frac{V_A}{I_C} = \frac{100 \text{ V}}{1 \text{ mA}} = 100 \text{ k}\Omega$$

$$C_\pi + C_\mu = \frac{g_m}{\omega_T} = \frac{40 \times 10^{-3}}{2\pi \times 800 \times 10^6} = 8 \text{ pF}$$

$$C_\mu = 1 \text{ pF} \qquad C_\pi = 7 \text{ pF}$$

The midband voltage gain is

$$A_M = -\left(\frac{r_\pi}{R_{sig} + r_\pi}\right)(g_m R_L')$$

where

$$R_L' = R_L \| r_o = 5 \| 100 = 4.76 \text{ k}\Omega$$

Thus,

$$g_m R_L' = 40 \times 4.76 = 190.5 \text{ V/V}$$

and

$$A_M = -\frac{2.5}{5 + 2.5} \times 190.5 = -63.5 \text{ V/V}$$

and

$$20\log|A_M| = 36 \text{ dB}$$

To determine f_H we first find C_{in},

$$C_{in} = C_\pi + C_\mu(1 + g_m R_L')$$
$$= 7 + 1(1 + 190.5) = 198.5 \text{ pF}$$

and the effective source resistance R_{sig}',

$$R_{sig}' = R_{sig} \| r_\pi = 5 \| 2.5 = 1.67 \text{ k}\Omega$$

Thus,

$$f_H = \frac{1}{2\pi C_{in} R_{sig}'} = \frac{1}{2\pi \times 198.5 \times 10^{-12} \times 1.67 \times 10^3} = 480 \text{kHz}$$

Finally, as in the case of the CS amplifier, it can be shown that the CE amplifier has a transmission zero with frequency

$$f_Z = \frac{g_m}{2\pi C_\mu} = \frac{40 \times 10^{-3}}{2\pi \times 1 \times 10^{-12}} = 6.37 \text{ GHz}$$

which is much higher than f_H.

EXERCISE

10.9 For the amplifier in Example 10.3, find the value of R_L that reduces the midband gain to half the value found. What value of f_H results? Note the trade-off between gain and bandwidth.
Ans. 2.44 kΩ; 923 MHz

10.2.5 Miller's Theorem

In our analysis of the high-frequency response of the common-source and common-emitter amplifiers, we employed a technique for replacing the bridging capacitance (C_{gd} or C_μ) by an equivalent input capacitance. This very useful and effective technique is based on a general theorem known as **Miller's theorem**, which we now present.

Consider the situation in Fig. 10.14(a). As part of a larger circuit that is not shown, we have isolated two circuit nodes, labeled 1 and 2, between which an impedance Z is connected. Nodes 1 and 2 are also connected to other parts of the circuit, as signified by the broken lines emanating from the two nodes. Furthermore, we assume that the voltage at node 2 is related to that at node 1 by

$$V_2 = KV_1 \tag{10.54}$$

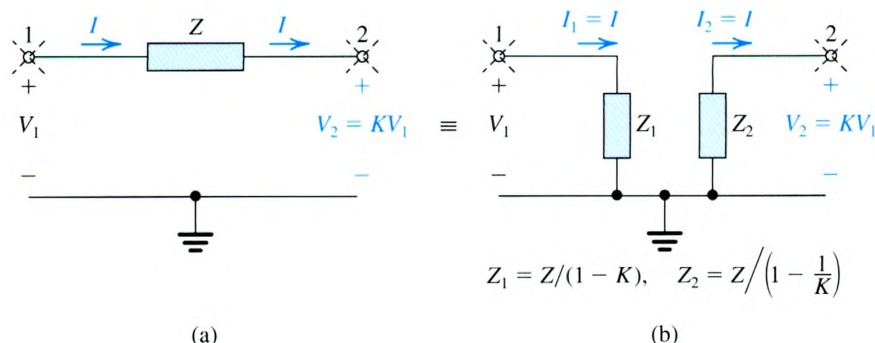


Figure 10.14 The Miller equivalent circuit.

In typical situations, K is a gain factor that can be positive or negative and has a magnitude usually larger than one.

Miller's theorem states that impedance Z can be replaced by two impedances: Z_1 connected between node 1 and ground and Z_2 connected between node 2 and ground, where

$$Z_1 = Z/(1-K) \tag{10.55a}$$

and

$$Z_2 = Z\Big/\left(1-\frac{1}{K}\right) \tag{10.55b}$$

to obtain the equivalent circuit shown in Fig. 10.14(b).

The proof of Miller's theorem is achieved by deriving Eqs. (10.55) as follows: In the original circuit of Fig. 10.14(a), the only way that node 1 "feels the existence" of impedance Z is through the current I that Z draws away from node 1. Therefore, to keep this current unchanged in the equivalent circuit, we must choose the value of Z_1 so that it draws an equal current,

$$I_1 = \frac{V_1}{Z_1} = I = \frac{V_1 - KV_1}{Z}$$

which yields the value of Z_1 in Eq. (10.55a). Similarly, to keep the current into node 2 unchanged, we must choose the value of Z_2 so that

$$I_2 = \frac{0 - V_2}{Z_2} = \frac{0 - KV_1}{Z_2} = I = \frac{V_1 - KV_1}{Z}$$

which yields the expression for Z_2 in Eq. (10.55b).

Example 10.4

Figure 10.15(a) shows an ideal voltage amplifier with a gain of −100 V/V and an impedance Z connected between its output and input terminals. Find the Miller equivalent circuit when Z is (a) a 1-MΩ resistance and (b) a 1-pF capacitance. In each case, use the equivalent circuit to determine V_o/V_{sig}.

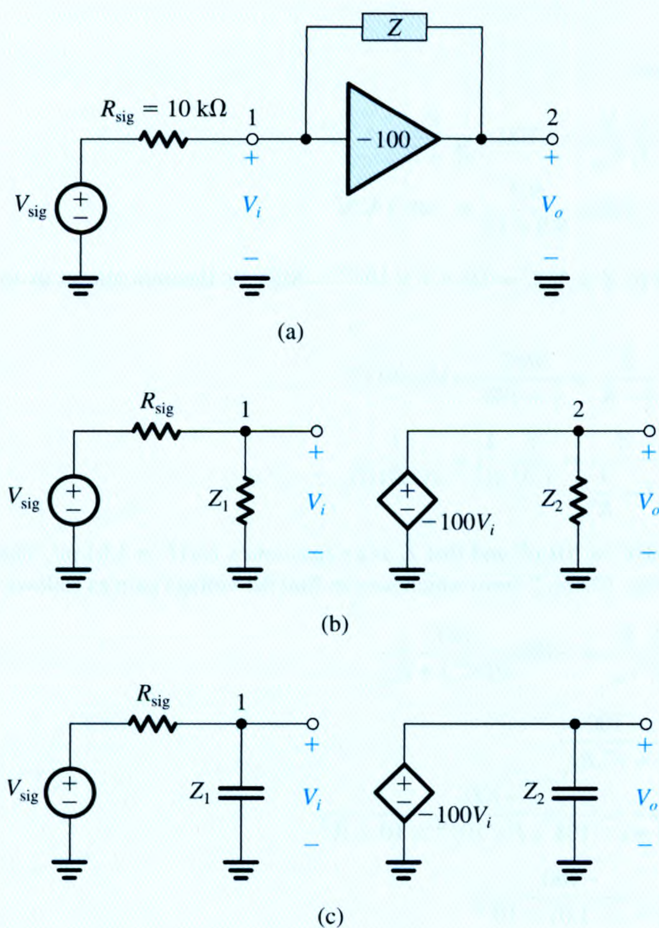


Figure 10.15 Circuits for Example 10.4.

Solution

(a) For $Z = 1\ \text{M}\Omega$, Miller's theorem gives us the equivalent circuit in Fig. 10.15(b), where

$$Z_1 = \frac{Z}{1-K} = \frac{1000\ \text{k}\Omega}{1+100} = 9.9\ \text{k}\Omega$$

$$Z_2 = \frac{Z}{1-\dfrac{1}{K}} = \frac{1\ \text{M}\Omega}{1+\dfrac{1}{100}} = 0.99\ \text{M}\Omega$$

Example 10.4 *continued*

The voltage gain can be found as follows:

$$\frac{V_o}{V_{sig}} = \frac{V_o}{V_i}\frac{V_i}{V_{sig}} = -100 \times \frac{Z_1}{Z_1 + R_{sig}}$$
$$= -100 \times \frac{9.9}{9.9+10} = -49.7 \text{ V/V}$$

(b) For Z as a 1-pF capacitance—that is, $Z = 1/sC = 1/s \times 1 \times 10^{-12}$ —Miller's theorem allows us to replace Z by Z_1 and Z_2, where

$$Z_1 = \frac{Z}{1-K} = \frac{1/sC}{1+100} = 1/s(101C)$$

$$Z_2 = \frac{Z}{1-\frac{1}{K}} = \frac{1}{1.01}\frac{1}{sC} = \frac{1}{s(1.01C)}$$

It follows that Z_1 is a capacitance $101C = 101$ pF and that Z_2 is a capacitance $1.01C = 1.01$ pF. The resulting equivalent circuit is shown in Fig. 10.15(c), from which we can find the voltage gain as follows:

$$\frac{V_o}{V_{sig}} = \frac{V_o}{V_i}\frac{V_i}{V_{sig}} = -100\frac{1/sC_1}{1/(sC_1)+R_{sig}}$$
$$= \frac{-100}{1+sC_1R_{sig}}$$
$$= \frac{-100}{1+s \times 101 \times 1 \times 10^{-12} \times 10 \times 10^3}$$
$$= \frac{-100}{1+s \times 1.01 \times 10^{-6}}$$

This is the transfer function of a first-order low-pass network with a dc gain of –100 V/V and a 3-dB frequency f_{3dB} of

$$f_{3dB} = \frac{1}{2\pi \times 1.01 \times 10^{-6}} = 157.6 \text{ kHz}$$

From Example 10.4, we can see that the Miller replacement of a feedback or bridging resistance results, for a negative K, in a smaller resistance [by a factor $(1-K)$] at the input. If the feedback element is a capacitance, its value is multiplied by $(1-K)$ to obtain the equivalent capacitance at the input side. The multiplication of a feedback capacitance by $(1-K)$ is called **Miller multiplication** or the **Miller effect**. We have encountered the Miller effect in the analysis of the CS and CE amplifiers. Note, however, that we neglected the Miller capacitance at the output because it is small; for the CS case, $C_2 = C_{gd}\left(1+\frac{1}{g_mR'_L}\right) \simeq C_{gd}$.

JOHN MILTON MILLER: CAPACITANCE MULTIPLICATION

In 1920 in the Scientific Papers of the National Bureau of Standards, John Miller first published data and analysis on the input capacitance of a vacuum-tube triode amplifier and the phenomenon ultimately known as the Miller effect. Attempts to eliminate or reduce this effect in amplifiers and oscillators led to several developments: First, in 1926, came vacuum-tube pentodes with internal grounded shielding elements; much later, in 1939, the cascode configuration was introduced, initially using vacuum-tube triodes and then BJTs and MOSFETs.

10.3 The Method of Open-Circuit Time Constants

In Section 10.2 we presented an approximate method, using the Miller effect, to analyze the high-frequency response of the CS and CE amplifiers; the method provides a reasonably accurate estimate of f_H and, equally important, considerable insight into the mechanism that limits high-frequency operation. Unfortunately, however, this method cannot deal with the case when a load capacitance C_L is present. As well, the method cannot easily be extended to more complex amplifier circuits. In the following we present an alternative method for estimating f_H. The method is both general and easy to use. Before doing so, however, we will digress briefly to consider the amplifier poles and zeros.

10.3.1 The High-Frequency Gain Function

The amplifier gain, taking into account the internal transistor capacitances, can be expressed as a function of the complex-frequency variable s in the general form

$$A(s) = A_M F_H(s) \tag{10.56}$$

where A_M is the midband gain, which for IC amplifiers is also equal to the low-frequency or dc gain (refer to Fig. 10.2). The value of A_M can be determined by analyzing the amplifier equivalent circuit while neglecting the effect of the transistor internal capacitances—in other words, by assuming that they act as perfect open circuits. By taking these capacitances into account, we see that the gain acquires the factor $F_H(s)$, which can be expressed in terms of its poles and zeros, which are usually real, as follows:

$$F_H(s) = \frac{(1+s/\omega_{Z1})(1+s/\omega_{Z2})\cdots(1+s/\omega_{Zn})}{(1+s/\omega_{P1})(1+s/\omega_{P2})\cdots(1+s/\omega_{Pn})} \tag{10.57}$$

where $\omega_{P1}, \omega_{P2}, \ldots, \omega_{Pn}$ are positive numbers representing the frequencies of the n real poles[2] and $\omega_{Z1}, \omega_{Z2}, \ldots, \omega_{Zn}$ are positive, negative, or infinite numbers representing the frequencies of the n real transmission zeros. Since the frequencies of the zeros and poles are by definition greater than the midband frequencies, we see from Eq. (10.57) that as s approaches midband frequencies, $F_H(s)$ approaches unity and the gain approaches A_M.

[2] In certain cases, some of the poles can be complex. One notable such case is the source and emitter followers, which can have a pair of complex-conjugate poles (Section 10.5).

10.3.2 Determining the 3-dB Frequency f_H

The amplifier designer usually is most interested in the part of the high-frequency band that is close to the midband. This is because the designer needs to estimate—and, if need be, modify—the value of the upper 3-dB frequency f_H (or ω_H; $f_H = \omega_H/2\pi$). In many cases the zeros are either at infinity or at such high frequencies as to be of little significance in determining ω_H. If in addition one of the poles, say ω_{P1}, is of much lower frequency than any of the other poles, then this pole will have the greatest effect on the value of the amplifier ω_H. In other words, this pole will *dominate* the high-frequency response of the amplifier, and the amplifier is said to have a **dominant-pole response**. In such cases, the function $F_H(s)$ can be approximated by

$$F_H(s) \simeq \frac{1}{1 + s/\omega_{P1}} \tag{10.58}$$

which is the transfer function of a first-order (or STC) low-pass network. It follows that if a dominant pole exists, then the determination of ω_H is greatly simplified:

$$\omega_H \simeq \omega_{P1} \tag{10.59}$$

This is the situation we encountered in the cases of the common-source and common-emitter amplifiers we analyzed in Section 10.2. As a rule of thumb, *a dominant pole exists if the lowest-frequency pole is at least two octaves (a factor of 4) away from the nearest pole or zero.*

EXERCISE

10.10 A direct-coupled amplifier has a dc gain of 1000 V/V and an upper 3-dB frequency of 100 kHz. Find the transfer function and the gain–bandwidth product in hertz.

Ans. $\dfrac{1000}{1 + \dfrac{s}{2\pi \times 10^5}}$; 10^8 Hz

In most practical cases, however, it is not easy to determine the poles and zeros. Nevertheless, we can obtain an estimate of ω_H using the method of open-circuit time constants, as follows.

Consider the function $F_H(s)$ [Eq. (10.57)], with the numerator and denominator factors multiplied out,

$$F_H(s) = \frac{1 + a_1 s + a_2 s^2 + \cdots + a_n s^n}{1 + b_1 s + b_2 s^2 + \cdots + b_n s^n} \tag{10.60}$$

where the coefficients a and b are related to the frequencies of the zeros and poles, respectively. Now, if a dominant pole exists, $F_H(s)$ can be approximated as

$$F_H(s) \simeq \frac{1}{1 + b_1 s} \tag{10.61}$$

and the 3-dB frequency ω_H is given by

$$\omega_H = \frac{1}{b_1}$$

The value of b_1 can be found by setting the input signal to zero and considering the various capacitances in the high-frequency equivalent circuit one at a time while reducing all other capacitors to zero (or, equivalently, replacing them with open circuits). That is, to obtain the contribution of capacitance C_i, we reduce all other capacitances to zero, reduce the input signal source to zero, and determine the resistance R_i seen by C_i. This can be done either by inspection or by replacing C_i with a voltage source V_x, finding the current I_x drawn from V_x, and calculating $R_i \equiv V_x/I_x$. The time constant τ_i associated with C_i is then found as $\tau_i = C_iR_i$. This process is then repeated for all other capacitors in the circuit. The value of b_1 is computed by summing the individual time constants, called **open-circuit time constants**,

$$b_1 = \sum_{i=1}^{n} C_iR_i \tag{10.62}$$

where we have assumed that there are n capacitors in the high-frequency equivalent circuit.

We will henceforth refer to b_1 as the **high-frequency time constant** and denote it τ_H. Thus, ω_H is found as

$$\omega_H = \frac{1}{b_1} = \frac{1}{\tau_H} = \frac{1}{\sum_i C_iR_i} \tag{10.63}$$

In complex circuits we usually do not know whether a dominant pole exists. Nevertheless, using Eq. (10.63) to determine ω_H normally yields remarkably good results[3] even if a dominant pole does not exist.

10.3.3 Applying the Method of Open-Circuit Time Constants to the CS Amplifier

Figure 10.16 shows the high-frequency equivalent circuit for the common-source amplifier. Here, resistance R_L' represents the total resistance between the output (drain) node and ground and includes r_o and R_L (if one is present). Similarly, C_L represents the total capacitance between the drain node and ground and includes the MOSFET's drain-to-body capacitance (C_{db}), the capacitance introduced by a current-source load, the input capacitance of a succeeding amplifier stage (if one is present), and in some cases, as we will see in later chapters, a deliberately introduced capacitance. In IC MOS amplifiers, C_L can be substantial.

We wish to determine the 3-dB frequency f_H of the CS amplifier in Fig. 10.16 using the method of open-circuit time constants. To that end, we set $V_{sig} = 0$ and consider each of the three capacitances at a time, setting the other two to zero. Figure 10.17(a) shows the resulting circuit for determining the resistance R_{gs} seen by C_{gs}, thus

$$R_{gs} = R_{sig} \tag{10.64}$$

For C_{gd}, we obtain the circuit in Fig. 10.17(b). This circuit is somewhat complex, and we cannot determine R_{gd} by inspection. Rather, we apply a test current source I_x and determine

---

[3] The method of open-circuit time constants yields good results only when all the poles are real, as is the case in most of this chapter.

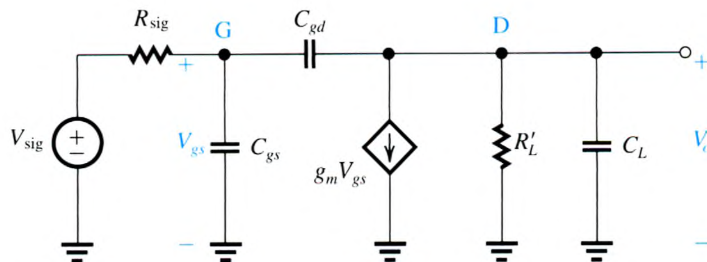


Figure 10.16 High-frequency equivalent circuit for the CS amplifier.

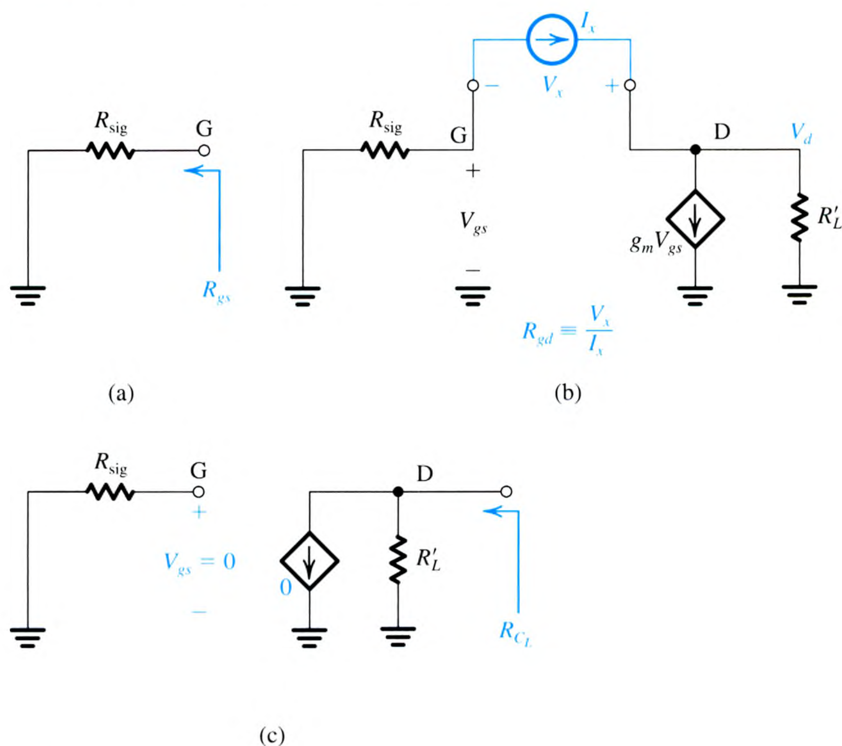


Figure 10.17 Application of the open-circuit time-constants method to the CS equivalent circuit of Fig. 10.16.

the voltage V_x that results across I_x. We see that

$$V_{gs} = -I_x R_{\text{sig}} \tag{10.65}$$

A loop equation gives

$$V_d = V_x + V_{gs}$$

A node equation at D gives

$$I_x = g_m V_{gs} + \frac{V_d}{R'_L}$$

Thus,

$$I_x = g_m V_{gs} + \frac{V_x + V_{gs}}{R'_L}$$

Substituting for V_{gs} from Eq. (10.65) and manipulating the resulting equation gives

$$R_{gd} \equiv \frac{V_x}{I_x} = R_{sig}(1 + g_m R'_L) + R'_L \tag{10.66}$$

Finally, for C_L we obtain the circuit shown in Fig. 10.17(c), from which

$$R_{C_L} = R'_L \tag{10.67}$$

Next, we use the resistance values in Eqs. (10.64), (10.66), and (10.67) to obtain the effective high-frequency time constant τ_H,

$$\begin{aligned}\tau_H &= C_{gs}R_{gs} + C_{gd}R_{gd} + C_L R_{C_L} \\ &= C_{gs}R_{sig} + C_{gd}[R_{sig}(1 + g_m R'_L) + R'_L] + C_L R'_L\end{aligned} \tag{10.68}$$

and the 3-dB frequency f_H is

$$f_H = \frac{1}{2\pi \tau_H}$$

An important observation is available from Eq. (10.68), which can be rewritten as

$$\tau_H = [C_{gs} + C_{gd}(1 + g_m R'_L)]R_{sig} + (C_{gd} + C_L)R'_L \tag{10.69}$$

We note that the first term is simply $C_{in}R_{sig}$, where C_{in} is dominated by the Miller capacitance $C_{gd}(1 + g_m R'_L)$. The method of open-circuit time constants, however, also provides the second term, which results from the interaction of $(C_{gd} + C_L)$ with R'_L. Thus, while the first term in Eq. (10.69) arises at the input node, the second term occurs at the output node. The second term can be dominant in cases where R_{sig} is small, as we have seen in Section 10.2.3.

Example 10.5

An integrated-circuit CS amplifier has $g_m = 2$ mA/V, $C_{gs} = 20$ fF, $C_{gd} = 5$ fF, $C_L = 25$ fF, $R_{sig} = 20\,\text{k}\Omega$, and $R'_L = 10\,\text{k}\Omega$. Determine f_H and the frequency of the transmission zero f_Z caused by C_{gd}. Note that this amplifier is the same as that analyzed in Example 10.1 except that here we include C_L.

Solution

To obtain f_H we first use Eqs. (10.64), (10.66), and (10.67) to determine the resistances seen by the three capacitors C_{gs}, C_{gd}, and C_L, respectively,

$$\begin{aligned}R_{gs} &= R_{sig} = 20\,\text{k}\Omega \\ R_{gd} &= R_{sig}(1 + g_m R'_L) + R'_L \\ &= 20(1 + 2 \times 10) + 10 = 430\,\text{k}\Omega \\ R_{C_L} &= R'_L = 10\,\text{k}\Omega\end{aligned}$$

Example 10.5 *continued*

We then determine the three time constants:

$$\tau_{gs} = C_{gs}R_{gs} = 20 \times 10^{-15} \times 20 \times 10^{3} = 400\,\text{ps}$$
$$\tau_{gd} = C_{gd}R_{gd} = 5 \times 10^{-15} \times 430 \times 10^{3} = 2150\,\text{ps}$$
$$\tau_{C_L} = C_L R_{C_L} = 25 \times 10^{-15} \times 10 \times 10^{3} = 250\,\text{ps}$$

The effective high-frequency time constant τ_H can now be obtained by summing the three time constants,

$$\tau_H = \tau_{gs} + \tau_{gd} + \tau_{C_L}$$
$$= 400 + 2150 + 250 = 2800\,\text{ps}$$

and the 3-dB frequency f_H is

$$f_H = \frac{1}{2\pi\,\tau_H} = \frac{1}{2\pi \times 2800 \times 10^{-12}} = 56.8\,\text{MHz}$$

which is lower than the estimate obtained in Example 10.1 using the Miller effect (63.7 MHz). This is because the method using the Miller effect does not allow for including the effect of C_L. Thus the lower estimate obtained by the method of open-circuit time constants is more realistic.

The frequency of the transmission zero can be determined by reference to the circuit in Fig. 10.16. Since at $s = s_Z$, V_o will be zero, a node equation at the output node at $s = s_Z$ becomes

$$s_Z C_{gd}(V_{gs} - 0) = g_m V_{gs}$$

Thus

$$s_Z = \frac{g_m}{C_{gd}}$$

and

$$f_Z = \frac{g_m}{2\pi\,C_{gd}} = \frac{2 \times 10^{-3}}{2\pi \times 5 \times 10^{-15}} = 63.7\,\text{GHz}$$

which is much higher than f_H; hence it plays almost no role in the estimate of f_H.

EXERCISES

10.11 For the CS amplifier in Example 10.5, using the value of f_H determined by the method of open-circuit time constants, find the gain–bandwidth product. Recall that $g_m = 2$ mA/V and $R_L' = 10\,\text{k}\Omega$.
Ans. GBW = 1.14 GHz

10.12 As a way to trade gain for bandwidth, the designer of the CS amplifier in Example 10.5 connects a load resistor at the output that results in halving the value of R_L'. Find the new values of $|A_M|$, f_H, and the gain–bandwidth product.
Ans. 10 V/V; 96.5 MHz; 0.965 GHz

10.13 As another way to trade dc gain for bandwidth, the designer of the CS amplifier in Example 10.5 decides to operate the amplifying transistor at double the value of V_{OV} by increasing the bias current fourfold. Find the new values of g_m, R_L', $|A_M|$, f_H, and the gain–bandwidth product. Assume that R_L' is the parallel equivalent of r_o of the amplifying transistor and that of the current-source load.

Ans. 4 mA/V; 2.5 kΩ; 10 V/V; 101 MHz; 1.01 GHz

We conclude this section by emphasizing that the method of open-circuit time constants reveals to the circuit designer the relative contribution of the various capacitances to the determination of the amplifier bandwidth f_H. For instance, for the amplifier in Example 10.5, we see that while C_{gd} contributes the most (2150 ps of 2800 pF, or 77%) because of the Miller effect, the contribution of each of C_{gs} (14%) and C_L (9%) is not insignificant. Such information can be useful in the amplifier-design process.

10.3.4 Application of the Method of Open-Circuit Time Constants to the CE Amplifier

The method of open-circuit time constants can be applied to the common-emitter amplifier to determine its 3-dB frequency in a manner analogous to that used in the case of the CS amplifier above. Specifically, consider the CE amplifier whose high-frequency equivalent circuit is shown in Fig. 10.13(b) but with the more realistic situation where a capacitance C_L is connected across R_L. Using Thévenin's theorem at the input, we can replace V_{sig}, R_{sig}, and r_π by

$$V_{sig}' = V_{sig}\frac{r_\pi}{r_\pi + R_{sig}}$$

and

$$R_{sig}' = R_{sig} \| r_\pi \tag{10.70}$$

The structure of the resulting circuit will be identical to that of the CS amplifier in Fig. 10.16 with V_{sig} replaced by V_{sig}', R_{sig} by R_{sig}', C_{gs} by C_π, and C_{gd} by C_μ. The formulas developed for the CS case can then be adapted for the CE case to obtain:

$$R_\pi = R_{sig}' \tag{10.71}$$

$$R_\mu = R_{sig}'(1 + g_m R_L') + R_L' \tag{10.72}$$

$$R_{C_L} = R_L' \tag{10.73}$$

Thus,

$$\tau_H = C_\pi R_{sig}' + C_\mu[R_{sig}'(1 + g_m R_L') + R_L'] + C_L R_L' \tag{10.74}$$

and

$$f_H = \frac{1}{2\pi\,\tau_H} \tag{10.75}$$

Video Example VE 10.2 Gain and Frequency Response of a CE Amplifier

A common-emitter amplifier has $C_\pi = 5$ pF, $C_\mu = 0.3$ pF, $C_L = 5$ pF, $g_m = 20$ mA/V, $\beta = 100$, $R'_L = 5$ kΩ, and $R_{sig} = 5$ kΩ. Find the midband gain A_M and an estimate of the 3-dB frequency f_H using the Miller effect. Also, obtain another estimate of f_H using the method of open-circuit time constants. Which of the two estimates would you consider to be more realistic, and why?


VE 10.2

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 10.48

EXERCISE

10.14 Consider a bipolar active-loaded CE amplifier having the load current source implemented with a *pnp* transistor. Let the circuit be operating at a 1-mA bias current. The transistors are specified as follows: $\beta(npn) = 200$, $V_{An} = 130$ V, $|V_{Ap}| = 50$ V, $C_\pi = 16$ pF, $C_\mu = 0.3$ pF, and $C_L = 5$ pF. The amplifier is fed with a signal source having a resistance of 36 kΩ. Determine: (a) A_M; (b) C_{in} and f_H using the Miller effect; (c) f_H using open-circuit time constants; (d) f_Z; and (e) the gain–bandwidth product.

Ans. (a) −176 V/V; (b) 450 pF, 80.6 kHz; (c) 73.5 kHz; (d) 21.2 GHz; (e) 12.9 MHz

10.4 High-Frequency Response of Common-Gate and Cascode Amplifiers

Although common-source and common-emitter amplifiers provide substantial gain at midband frequencies, their gain falls off in the high-frequency band at a relatively low frequency. This is primarily due to the large input capacitance C_{in}, whose value is significantly increased by the Miller component. The latter is large because of the Miller multiplication effect that the bridging capacitance C_{gd} (or C_μ) experiences. It follows that the key to obtaining wideband operation, that is, high f_H, is to use circuit configurations that do not suffer from the Miller effect. One such configuration is the common-gate circuit.

10.4.1 High-Frequency Response of the CG Amplifier

Figure 10.18(a) shows a CG amplifier with the MOSFET internal capacitances C_{gs} and C_{gd} pulled out of the model and indicated. For generality, a capacitance C_L is included at the output node to represent the combination of the output capacitance of a current-source load and the input capacitance of a succeeding amplifier stage. Capacitance C_L also includes the

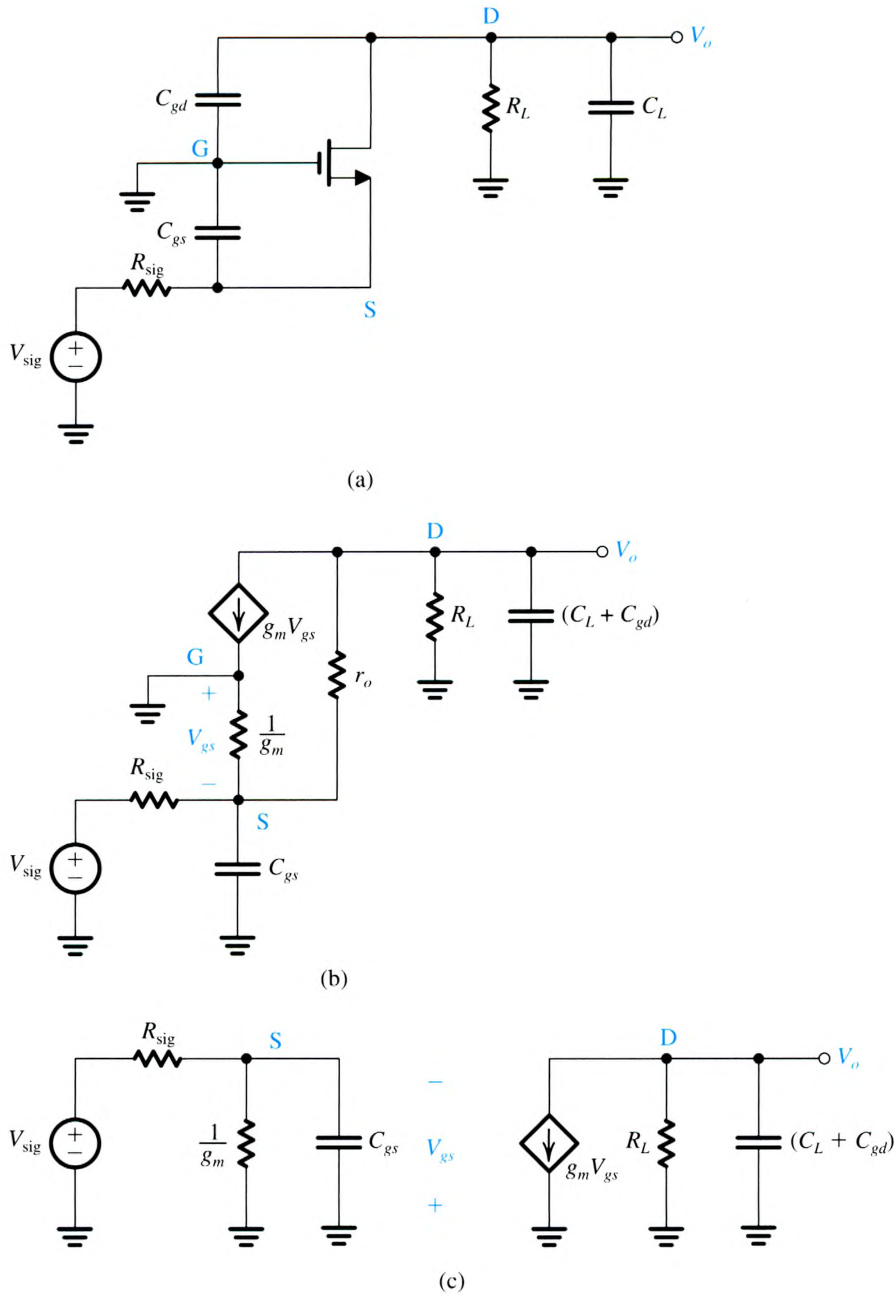


Figure 10.18 **(a)** The common-gate amplifier with the transistor internal capacitances shown. A load capacitance C_L is also included. **(b)** Equivalent circuit of the CG amplifier with the MOSFET replaced with its T model. **(c)** Equivalent circuit for the case in which r_o is neglected. **(d)** Circuit for determining the resistance R_{gs} seen by C_{gs}. **(e)** Circuit for determining the resistance R_{gd} seen by $(C_L + C_{gd})$.

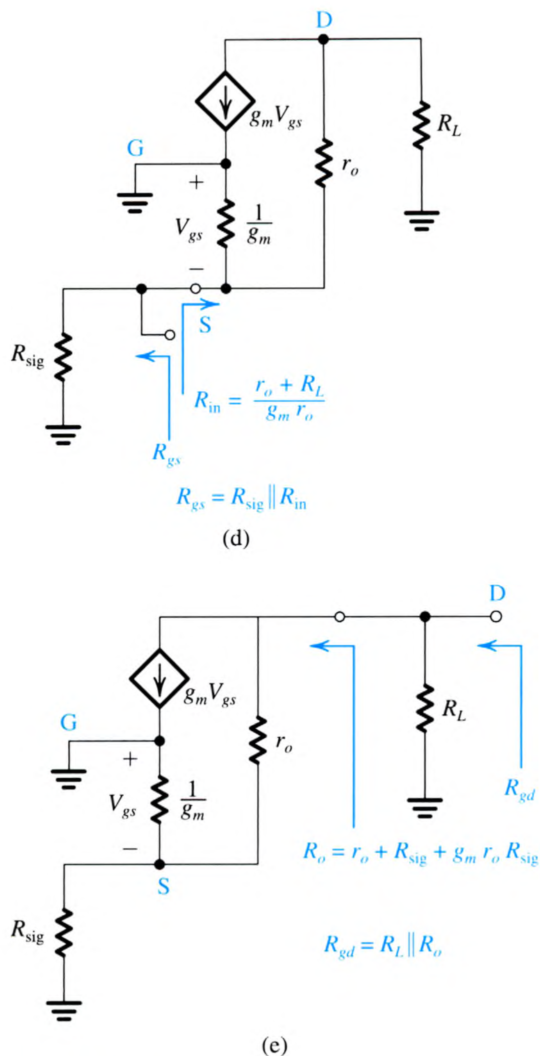


Figure 10.18 *continued*

MOSFET capacitance C_{db}. Note the C_L appears in effect in parallel with C_{gd}; therefore, in the following analysis we will lump the two capacitances together.

It is important to note at the outset that each of the three capacitances in the circuit of Fig. 10.18(a) has a grounded node. Therefore *none of the capacitances undergoes the Miller multiplication effect* observed in the CS stage. It follows that the CG circuit can be designed to have a much wider bandwidth than that of the CS circuit, especially when the resistance of the signal generator is large. To analyze the high-frequency response of the CG amplifier of Fig. 10.18(a), we replace the MOSFET with its T model. The resulting circuit, with C_{gd} lumped with C_L, is shown in Fig. 10.18(b).

We shall consider first the case of a discrete-circuit CG amplifier in which r_o can be neglected. Eliminating r_o results in the circuit in Fig. 10.18(c). We immediately observe that

there are two poles: one at the input side with a frequency f_{P1},

$$f_{P1} = \frac{1}{2\pi C_{gs}\left(R_{sig} \parallel \dfrac{1}{g_m}\right)} \tag{10.76}$$

and the other at the output side with a frequency f_{P2},

$$f_{P2} = \frac{1}{2\pi(C_{gd} + C_L)R_L} \tag{10.77}$$

The relative locations of the two poles will depend on the specific situation. However, f_{P2} is usually lower than f_{P1}; thus f_{P2} can be dominant. The important point to note is that both f_{P1} and f_{P2} are usually much higher than the frequency of the dominant input pole in the CS stage. An approximate value for f_H can be obtained by applying the method of open-circuit time constants to the circuit of Fig. 10.18(c), resulting in

$$\tau_{gs} = C_{gs}\left(R_{sig} \parallel \frac{1}{g_m}\right) = 1/2\pi f_{P1} \tag{10.78}$$

and

$$\tau_{gd} = (C_L + C_{gd})R_L = 1/2\pi f_{P2} \tag{10.79}$$

Thus,

$$\tau_H = C_{gs}\left(R_{sig} \parallel \frac{1}{g_m}\right) + (C_L + C_{gd})R_L \tag{10.80}$$

and

$$f_H = \frac{1}{2\pi \tau_H} = 1\bigg/\left(\frac{1}{f_{P1}} + \frac{1}{f_{P2}}\right) \tag{10.81}$$

In IC amplifiers, r_o has to be taken into account. Applying the method of open-circuit time constants to the equivalent circuit in Fig. 10.18(b), we obtain the circuit in Fig. 10.18(d) for determining R_{gs}. From this circuit we find that

$$R_{gs} = R_{sig} \parallel R_{in} \tag{10.82}$$

where R_{in} is the input resistance of the CG amplifier with a load resistance R_L. An expression for R_{in} was derived in Chapter 8 and given in Eq. (8.51),

$$R_{in} = \frac{r_o + R_L}{1 + g_m r_o} \simeq \frac{r_o + R_L}{g_m r_o} \tag{10.83}$$

The resistance R_{gd} seen by $(C_L + C_{gd})$ can be obtained from the circuit in Fig. 10.18(e),

$$R_{gd} = R_L \parallel R_o \tag{10.84}$$

where R_o is the output resistance of a CG amplifier with a resistance R_{sig} connected between source and ground. From Chapter 8, Eq. (8.55), we have

$$R_o = r_o + R_{sig} + g_m r_o R_{sig} \tag{10.85}$$

Finally,

$$\tau_H = \tau_{gs} + \tau_{gd} \tag{10.86}$$

and

$$f_H = \frac{1}{2\pi \tau_H} \tag{10.87}$$

Example 10.6

Consider a common-gate amplifier with $g_m = 2$ mA/V, $r_o = 20$ kΩ, $C_{gs} = 20$ fF, $C_{gd} = 5$ fF, $C_L = 25$ fF, $R_{sig} = 20$ kΩ, and $R_L = 20$ kΩ. Assume that C_L includes C_{db}. Determine the input resistance, the midband gain, and the upper 3-dB frequency f_H.

Solution

Figure 10.19 shows the CG amplifier circuit at midband frequencies. We note that

$$v_o = iR_L$$

$$v_{sig} = i(R_{sig} + R_{in})$$

Thus, the overall voltage gain is given by

$$G_v = \frac{v_o}{v_{sig}} = \frac{R_L}{R_{sig} + R_{in}}$$

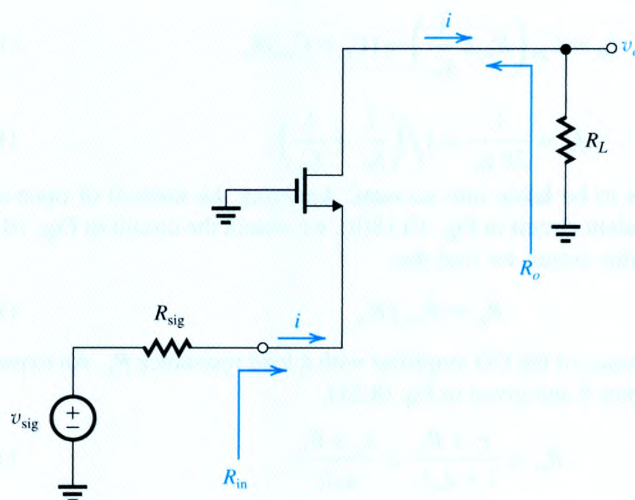


Figure 10.19 The CG amplifier circuit at midband.

The value of R_{in} is found from Eq. (10.83) as

$$R_{in} = \frac{r_o + R_L}{1 + g_m r_o}$$

$$= \frac{20 + 20}{1 + (2 \times 20)} = 0.98 \text{ k}\Omega$$

Thus, G_v can now be determined as

$$G_v = \frac{20}{20+0.98} = 0.95\text{ V/V}$$

Observe that as expected G_v is very low. This is due to the fact that the CG amplifier draws a large input current, equal, in fact, to the load current i.

To obtain an estimate of the 3-dB frequency f_H, we first determine R_{gs} and R_{gd} using Eqs. (10.82) and (10.84),

$$R_{gs} = R_{\text{sig}} \| R_{\text{in}} = 20 \| 0.98 = 0.93\text{ k}\Omega$$

$$R_{gd} = R_L \| R_o$$

where R_o is given by Eq. (10.85),

$$R_o = r_o + R_{\text{sig}} + (g_m r_o)R_{\text{sig}}$$

$$= 20 + 20 + 40 \times 20 = 840\text{ k}\Omega$$

Thus,

$$R_{gd} = 20 \| 840 = 19.5\text{ k}\Omega$$

Now we can compute the sum of the open-circuit time constants, τ_H,

$$\tau_H = C_{gs}R_{gs} + (C_{gd} + C_L)R_{gd}$$

$$\tau_H = 20 \times 10^{-15} \times 0.93 \times 10^3 + (5+25) \times 10^{-15} \times 19.5 \times 10^3$$

$$= 18.6 \times 10^{-12} + 585 \times 10^{-12}$$

$$= 603.6\text{ ps}$$

and the upper 3-dB frequency f_H can be obtained as

$$f_H = \frac{1}{2\pi\tau_H} = \frac{1}{2\pi \times 603.6 \times 10^{-12}} = 263.7\text{ MHz}$$

Observe that f_H is indeed much higher than (about five times) the corresponding value for the CS amplifier found in Example 10.5. Another important observation can be made by examining the two components of τ_H: The contribution of the input circuit is 18.6 ps, while that of the output circuit is 585 ps; thus the limitation on the high-frequency response is posed by the output circuit.

EXERCISE

10.15 In order to raise the midband gain of the CG amplifier in Example 10.6, the circuit designer decides to use a cascode current source for the load device, thus raising R_L by a factor of $g_m r_o = 40$; that is, R_L becomes 800 kΩ. Find R_{in}, the midband gain, and f_H. Comment on the results.

Ans. 20 kΩ; 20 V/V; 12.7 MHz. While the midband gain has been increased substantially (by a factor of 21), the bandwidth f_H has been substantially lowered (by a factor of about 20.7). Thus, the high-frequency advantage of the CG amplifier is completely lost!

We conclude this section by noting that a properly designed CG circuit can have a wide bandwidth. However, the input resistance will be low and the overall midband gain can be very low. It follows that the CG circuit alone will not do the job! However, combining the CG with the CS amplifier in the cascode configuration can result in a circuit having the high input resistance and gain of the CS amplifier together with the wide bandwidth of the CG amplifier, as we shall now see.

10.4.2 High-Frequency Response of the MOS Cascode Amplifier

In Section 8.5 we studied the cascode amplifier and analyzed its performance at midband frequencies. There we learned that by combining the CS and CG configurations, the cascode amplifier exhibits a very high input resistance and a voltage gain that can be as high as A_0^2, where $A_0 = g_m r_o$ is the intrinsic gain of the MOSFET. For our purposes here, we will see that the versatility of the cascode circuit allows us to trade off some of this high midband gain in return for a wider bandwidth.

Figure 10.20 shows the cascode amplifier with all transistor internal capacitances indicated. Also included is a capacitance C_L at the output node to represent the combination

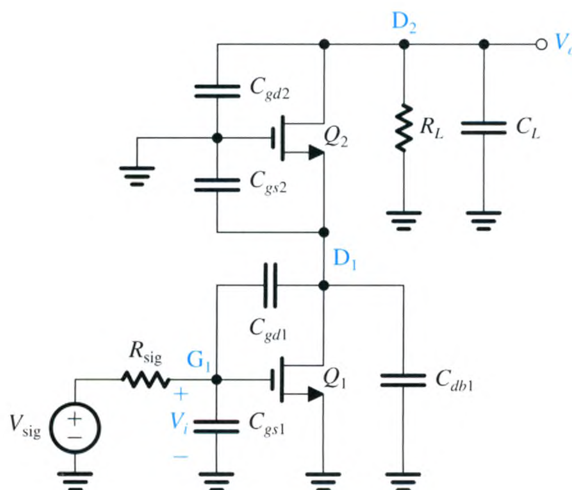


Figure 10.20 The cascode circuit with the various transistor capacitances indicated.

of C_{db2}, the output capacitance of a current-source load, and the input capacitance of a succeeding amplifier stage (if any). Note that C_{db1} and C_{gs2} appear in parallel, and we will combine them in the following analysis. Similarly, C_L and C_{gd2} appear in parallel and will be combined.

The easiest and, in fact, quite insightful approach to determining the 3-dB frequency f_H is to employ the open-circuit time-constants method:

1. Capacitance C_{gs1} sees a resistance R_{sig}.
2. Capacitance C_{gd1} is the gate-to-drain capacitance of the CS amplifier Q_1; thus it sees a resistance R_{gd1}, which can be obtained by adapting the formula in Eq. (10.66),

$$R_{gd1} = (1 + g_{m1}R_{d1})R_{sig} + R_{d1} \tag{10.88}$$

where R_{d1}, the total resistance at D_1, is given by the parallel equivalent of the resistance looking into the drain of Q_1 (r_{o1}) and the resistance looking into the source of Q_2 (R_{in2}), thus

$$R_{d1} = r_{o1} \parallel R_{in2} = r_{o1} \parallel \frac{r_{o2} + R_L}{g_{m2}r_{o2}} \tag{10.89}$$

3. Capacitance $(C_{db1} + C_{gs2})$ sees a resistance R_{d1}.
4. Capacitance $(C_L + C_{gd2})$ sees a resistance $(R_L \parallel R_o)$ where R_o, the output resistance of the cascode amplifier, is given by

$$R_o = r_{o2} + r_{o1} + (g_{m2}r_{o2})r_{o1} \tag{10.90}$$

With the resistances determined, the effective time constant τ_H can be computed as

$$\begin{aligned}\tau_H = {} & C_{gs1}R_{sig} + C_{gd1}\left[(1 + g_{m1}R_{d1})R_{sig} + R_{d1}\right] \\ & + \left(C_{db1} + C_{gs2}\right)R_{d1} + \left(C_L + C_{gd2}\right)(R_L \parallel R_o)\end{aligned} \tag{10.91}$$

and the 3-dB frequency f_H as

$$f_H = \frac{1}{2\pi\tau_H}$$

Design Insight and Trade-Offs To better understand what limits the high-frequency gain of the MOS cascode amplifier, we rewrite Eq. (10.91) in the form

$$\begin{aligned}\tau_H = {} & R_{sig}\left[C_{gs1} + C_{gd1}(1 + g_{m1}R_{d1})\right] + R_{d1}\left(C_{gd1} + C_{db1} + C_{gs2}\right) \\ & + (R_L \parallel R_o)\left(C_L + C_{gd2}\right)\end{aligned} \tag{10.92}$$

The first term arises at the input node, the second term at the middle node, namely (D_1, S_2), and the third term at the output node. The first term is simply due to the interaction of the signal-source resistance R_{sig} and the input capacitance of Q_1, which, as expected, includes the Miller capacitance $C_{gd1}(1 + g_{m1}R_{d1})$.

In the case of a large R_{sig}, the first term can dominate, especially if the Miller multiplier $(1 + g_{m1}R_{d1})$ is large. This in turn happens when the load resistance R_L is large (on the order of A_0r_o), causing R_{in2} and hence R_{d1} to be large and requiring the first stage, Q_1, to provide a large proportion of the gain (see Section 8.5.2). It follows that when R_{sig} is large, to extend the bandwidth we have to lower R_L to the order of r_o. This in turn lowers R_{in2} and hence

R_{d1} and renders the Miller effect in Q_1 insignificant. Note, however, that the dc gain of the cascode will then be A_0. Thus, while the dc gain will be the same as (or a little higher than) that achieved in a CS amplifier, the bandwidth will be greater.

In the case when R_{sig} is small, the Miller effect in Q_1 will not be of concern. A large value of R_L (on the order of $A_0 r_o$) can then be used to realize the large dc gain possible with a cascode amplifier—that is, a dc gain on the order of A_0^2. Equation (10.92) indicates that in this case the third term will usually be dominant. To pursue this point a little further, consider the case $R_{sig} = 0$, and assume that the middle term is much smaller than the third term. It follows that

$$\tau_H \simeq (C_L + C_{gd2})(R_L \| R_o) \tag{10.93}$$

	Common Source	Cascode
Circuit	V_i, r_o, R_L, C_L, V_o; $R'_L = R_L \| r_o$	V_i, $A_0 r_o$, $A_0 R_L$, C_L, V_o
DC Gain	$-g_m R'_L$	$-A_0 g_m R'_L$
f_{3dB}	$\dfrac{1}{2\pi(C_L + C_{gd})R'_L}$	$\dfrac{1}{2\pi(C_L + C_{gd})A_0 R'_L}$
f_t	$\dfrac{g_m}{2\pi(C_L + C_{gd})}$	$\dfrac{g_m}{2\pi(C_L + C_{gd})}$

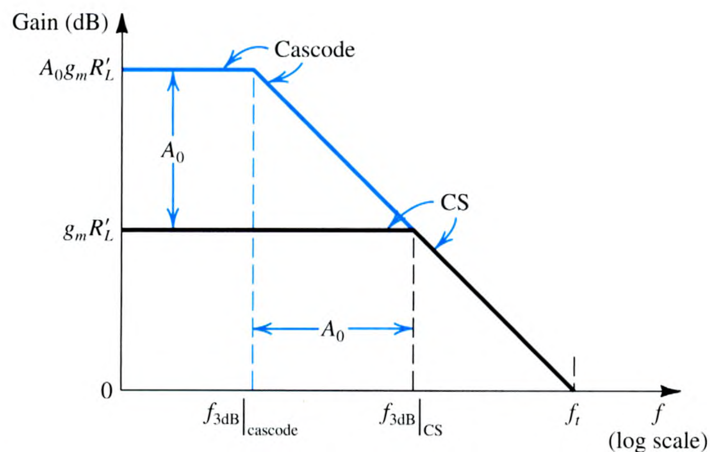


Figure 10.21 Effect of cascoding on gain and bandwidth in the case $R_{sig} = 0$. Cascoding can increase the dc gain by a factor equal to the intrinsic gain A_0 while keeping the unity-gain frequency constant. Note that to achieve the high gain, the load resistance must be increased by the factor A_0.

and the 3-dB frequency becomes

$$f_H = \frac{1}{2\pi(C_L + C_{gd2})(R_L \| R_o)} \tag{10.94}$$

which is of the same form as the formula for the CS amplifier with $R_{sig} = 0$ (Eq. 10.45). Here, however, $(R_L \| R_o)$ is larger than R'_L by a factor of about A_0. Thus f_H of the cascode will be lower than that of the CS amplifier by the same factor A_0. Figure 10.21 shows a sketch of the frequency response of the cascode and of the corresponding common-source amplifier. We observe that in this case, cascoding increases the dc gain by a factor A_0 while keeping the unity-gain frequency unchanged at

$$f_t \simeq \frac{1}{2\pi}\frac{g_m}{C_L + C_{gd2}} \tag{10.95}$$

Example 10.7

This example illustrates the advantages of cascoding by comparing the performance of a cascode amplifier with that of a common-source amplifier in two cases:

(a) The resistance of the signal source is significant, $R_{sig} = 20\text{ k}\Omega$.
(b) R_{sig} is negligibly small.

Assume all MOSFETs have $g_m = 2\text{ mA/V}$, $r_o = 20\text{ k}\Omega$, $C_{gs} = 20\text{ fF}$, $C_{gd} = 5\text{ fF}$, $C_{db} = 5\text{ fF}$, and C_L (excluding C_{db}) = 20 fF. For case (a), let $R_L = r_o = 20\text{ k}\Omega$ for both amplifiers. For case (b), let $R_L = r_o = 20\text{ k}\Omega$ for the CS amplifier and $R_L = R_o$ for the cascode amplifier. For all cases, determine A_v, f_H, and f_t.

Solution

(a) For the CS amplifier:

$$A_0 = g_m r_o = 2 \times 20 = 40\text{ V/V}$$
$$A_v = -g_m(R_L \| r_o) = -g_m(r_o \| r_o)$$
$$= -\frac{1}{2}A_0 = -20\text{ V/V}$$

To obtain τ_H we use Eq. (10.68) and note that here C_L does not include C_{db}, thus

$$\tau_H = C_{gs}R_{sig} + C_{gd}[(1 + g_m R'_L)R_{sig} + R'_L] + (C_L + C_{db})R'_L$$

where

$$R'_L = r_o \| R_L = r_o \| r_o = 10\text{ k}\Omega$$
$$\tau_H = 20 \times 20 + 5[(1 + 20)20 + 10] + (20 + 5)10$$
$$= 400 + 2150 + 250 = 2800\text{ ps}$$

Thus,

$$f_H = \frac{1}{2\pi \times 2800 \times 10^{-12}} = 56.8\text{ MHz}$$
$$f_t = |A_v| f_H = 20 \times 56.8 = 1.14\text{ GHz}$$

Example 10.7 *continued*

For the cascode amplifier:

$$R_o = 2r_o + (g_m r_o)r_o = (2 \times 20) + (40 \times 20) = 840\ \text{k}\Omega$$
$$A_v = -g_m(R_o \| R_L)$$
$$= -2(840 \| 20) = -39\ \text{V/V}$$
$$R_{\text{in2}} = \frac{r_o + R_L}{g_m r_o} = \frac{r_o + r_o}{g_m r_o} = \frac{2}{g_m} = \frac{2}{2} = 1.0\ \text{k}\Omega$$
$$R_{d1} = r_o \| R_{\text{in2}} = 20 \| 1 = 0.95\ \text{k}\Omega$$

Using Eq. (10.92),

$$\tau_H = R_{\text{sig}}[C_{gs1} + C_{gd1}(1 + g_{m1}R_{d1})]$$
$$+ R_{d1}(C_{gd1} + C_{db1} + C_{gs2})$$
$$+ (R_L \| R_o)(C_L + C_{db2} + C_{gd2})$$
$$= 20[20 + 5(1 + 2 \times 0.95)]$$
$$+ 0.95(5 + 5 + 20)$$
$$+ (20 \| 840)(20 + 5 + 5)$$
$$= 690 + 28.5 + 586$$
$$= 1305\ \text{ps}$$
$$f_H = \frac{1}{2\pi \times 1305 \times 10^{-12}} = 122\ \text{MHz}$$
$$f_t = 39 \times 122 = 4.76\ \text{GHz}$$

Thus cascoding has increased both the dc gain and the 3-dB frequency, with the combined effect being an increase of f_t by a factor of 4.2.

(b) For the CS amplifier:

$$A_v = -20\ \text{V/V}$$
$$\tau_H = (C_{gd} + C_L + C_{db})R'_L$$
$$= (5 + 20 + 5)10 = 300\ \text{ps}$$
$$f_H = \frac{1}{2\pi \times 300 \times 10^{-12}} = 530.5\ \text{MHz}$$
$$f_t = 20 \times 530.5 = 10.6\ \text{GHz}$$

For the cascode amplifier:

$$R_L = R_o = 840 \text{ k}\Omega$$

$$A_v = -g_m(R_o \parallel R_L)$$

$$= -2(840 \parallel 840) = -840 \text{ V/V}$$

$$R_{\text{in2}} = \frac{r_o + R_L}{g_m r_o} = \frac{20 + 840}{2 \times 20} = 21.5 \text{ k}\Omega$$

$$R_{d1} = r_{o1} \parallel R_{\text{in2}} = 20 \parallel 21.5 = 10.4 \text{ k}\Omega$$

$$\tau_H = R_{d1}(C_{gd1} + C_{db1} + C_{gs2}) + (R_L \parallel R_o)(C_L + C_{gd2} + C_{db2})$$

$$= 10.4(5 + 5 + 20) + (840 \parallel 840)(20 + 5 + 5)$$

$$= 312 + 12{,}600 = 12{,}912 \text{ ps}$$

$$f_H = \frac{1}{2\pi \times 12{,}912 \times 10^{-12}} = 12.3 \text{ MHz}$$

$$f_t = 840 \times 12.3 = 10.33 \text{ GHz}$$

Thus cascoding increases the dc gain from 20 V/V to 840 V/V; that is, by a factor of 42. This increase has been obtained at the cost of a decrease in f_H by approximately the same factor, resulting in the unity-gain frequency (which, in this case, is equal to the gain–bandwidth product) remaining nearly constant.

Video Example VE 10.3 Gain and High-Frequency Response of a MOS Cascode Amplifier

Find the dc gain and the 3-dB frequency of a MOS cascode amplifier operated at $g_m = 3$ mA/V and $r_o = 10$ kΩ. The MOSFETs have $C_{gs} = 15$ fF, $C_{gd} = 4$ fF, and $C_{db} = 4$ fF. The amplifier is fed from a signal source with $R_{\text{sig}} = 100$ kΩ and is connected to a load resistance of 320 kΩ. There is also a load capacitance C_L of 15 fF.

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.


VE 10.3

Related end-of-chapter problem: **10.56**

Video Example VE 10.4 Design and Frequency-Response Analysis of a MOS Cascode Amplifier

Design a cascode amplifer to provide a dc gain of 74 dB when driven with a low-resistance generator and utilizing NMOS transistors for which $V_A = 10$ V, $\mu_n C_{ox} = 400\ \mu\text{A/V}^2$, $W/L = 50$, $C_{gd} = 5$ fF, and $C_L = 50$ fF. Assuming that $R_L = R_o$, determine the overdrive voltage and the drain current at which the MOSFETs should be operated. Find the 3-dB frequency and the unity-gain frequency. If the cascode transistor is removed and R_L remains unchanged, what will the dc gain, the 3-dB frequency, and the unity-gain frequency become?


VE 10.4

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: **10.58**

EXERCISE

10.16 In this exercise we wish to contrast the gain and bandwidth of a CS amplifier and a cascode amplifier. Assume that both are fed with a large source resistance R_{sig} that effectively determines the high-frequency response. Thus, neglect components of τ_H that do not include R_{sig}. Also assume that all transistors are operated at the same conditions and thus corresponding small-signal parameters are equal. Also, both amplifiers have equal $R_L = r_o$, and $g_m r_o = 40$.

(a) Find the ratio of the low-frequency gain of the cascode amplifier to that of the CS amplifier.

(b) For the case of $C_{gd} = 0.25C_{gs}$, find the ratio of f_H of the cascode to that of the CS amplifier.

(c) Use (a) and (b) to find the ratio of f_t of the cascode to that of the CS.

Ans. 2; 3.6; 7.2

Conclusion The MOS cascode amplifier is a versatile circuit that, depending on the application at hand, can be designed to provide either higher dc gain or larger bandwidth than the CS amplifier or a combination of both.

10.4.3 High-Frequency Response of the Bipolar Cascode Amplifier

The analysis method studied in the previous section can be directly applied to the BJT cascode amplifier. Figure 10.22 presents the circuits and the formulas for determining the

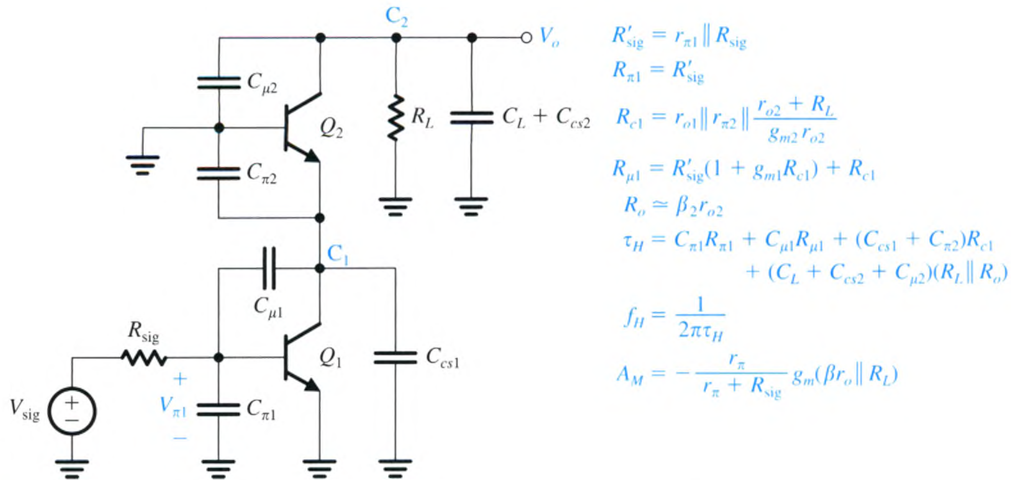


Figure 10.22 Determining the frequency response of the BJT cascode amplifier. Note that in addition to the BJT capacitances C_π and C_μ, the capacitances between the collector and the substrate C_{cs} for each transistor are included.

high-frequency response of the bipolar cascode. Note that some of these formulas rely on the study of the bipolar cascode in Section 8.5.3.

EXERCISE

10.17 The objective of this exercise is to evaluate the effect of cascoding on the performance of the CE amplifier of Exercise 10.14. The specifications are as follows: $I = 1$ mA, $\beta = 200$, $r_o = 130$ kΩ, $C_\pi = 16$ pF, $C_\mu = 0.3$ pF, $C_{cs1} = C_{cs2} = 0$, $C_L = 5$ pF, $R_{sig} = 36$ kΩ, $R_L = 50$ kΩ. Find R_{in}, A_0, R_{o1}, R_{in2}, R_o, A_M, f_H, and the gain–bandwidth product. Compare A_M, f_H, and GB with the corresponding values obtained in Exercise 10.14 for the CE amplifier. What should C_L be reduced to in order to have $f_H = 1$ MHz?

Ans. 5 kΩ; 5200 V/V; 130 kΩ; 35 Ω; 26 MΩ; −244 V/V; 470 kHz; 114.7 MHz. $|A_M|$ has increased from 176 V/V to 244 V/V; f_H has increased from 73.5 kHz to 470 kHz; GB has increased from 12.9 MHz to 114.7 MHz. C_L must be reduced to 1.41 pF.

10.5 High-Frequency Response of Source and Emitter Followers



In this section, we study the high-frequency response of two important circuit building blocks: the source follower and the emitter follower. Both have a midband voltage gain that is less than but close to unity. Their advantage lies in their high input resistance and low output resistance which make them useful as the output stage of a multistage amplifier and

as voltage buffers. As we will see, these voltage followers have another important advantage: a wide bandwidth.

Analyzing the high-frequency response of the source and emitter followers is complicated because the follower has two high-frequency poles that can be close to each other on the negative real axis of the s plane. Furthermore, in many cases, the poles can become complex. As a result, the method of open-circuit time constants does not provide a good estimate of f_H of the followers except in special circumstances. Our approach, therefore, will be to analyze the follower circuit to determine its gain V_o/V_{sig} as a function of frequency and then use it to determine f_H. Although the analysis is somewhat lengthy, the results can be applied easily. In the following we shall do the analysis of the source follower in detail. Then, because of their similarity, we will give the results for the emitter follower without proof.

10.5.1 The Source-Follower Case

Figure 10.23(a) shows a source follower without the biasing arrangement. The follower is driven by a signal source (V_{sig}, R_{sig}) and is loaded with a resistance R_L and, for generality, a capacitance C_L. Replacing the MOSFET with its hybrid-π equivalent-circuit model results in the equivalent circuit shown in Fig. 10.23(b). Here, we have included the body-effect generator $g_{mb}V_{bs}$ because it plays an important role in determining the source-follower gain. Also, we are assuming that whatever capacitances exist between the MOSFET source and ground, such as C_{sb} of Fig. 10.4(a), have been lumped into C_L.

Noting that the drain terminal is grounded, we see that C_{gd} in fact appears across the input terminals of the source follower. Also, r_o is in parallel with R_L and can be combined with it. Finally, since the body terminal B is connected to ground, the voltage V_{bs} appears across the controlled source $g_{mb}V_{bs}$. Thus we can use the source-absorption theorem (see Appendix D) to replace the controlled source with a resistance $1/g_{mb}$. Since the latter appears between source and ground, it is in parallel with R_L and can be combined with it.

Utilizing the above observations, we obtain the simplified equivalent circuit shown in Fig. 10.23(c) where

$$R_L' = R_L \,\|\, r_o \,\|\, \frac{1}{g_{mb}} \tag{10.96}$$

Obtaining the Transfer Function $V_o(s)/V_{sig}(s)$ Analysis of the equivalent circuit in Fig. 10.23(c) to determine the transfer function $V_o(s)/V_{sig}(s)$ proceeds as follows.

1. A loop equation at the input yields

$$V_{sig} = I_i R_{sig} + V_g$$

where V_g can be expressed as

$$V_g = V_{gs} + V_o$$

Thus,

$$V_{sig} = I_i R_{sig} + V_{gs} + V_o \tag{10.97}$$

2. A node equation at G provides

$$\begin{aligned} I_i &= sC_{gd}V_g + sC_{gs}V_{gs} \\ &= sC_{gd}(V_{gs} + V_o) + sC_{gs}V_{gs} \end{aligned}$$

which can be substituted into Eq. (10.97) to obtain

$$V_{sig} = [1 + s(C_{gs} + C_{gd})R_{sig}]V_{gs} + [1 + sC_{gd}R_{sig}]V_o \tag{10.98}$$

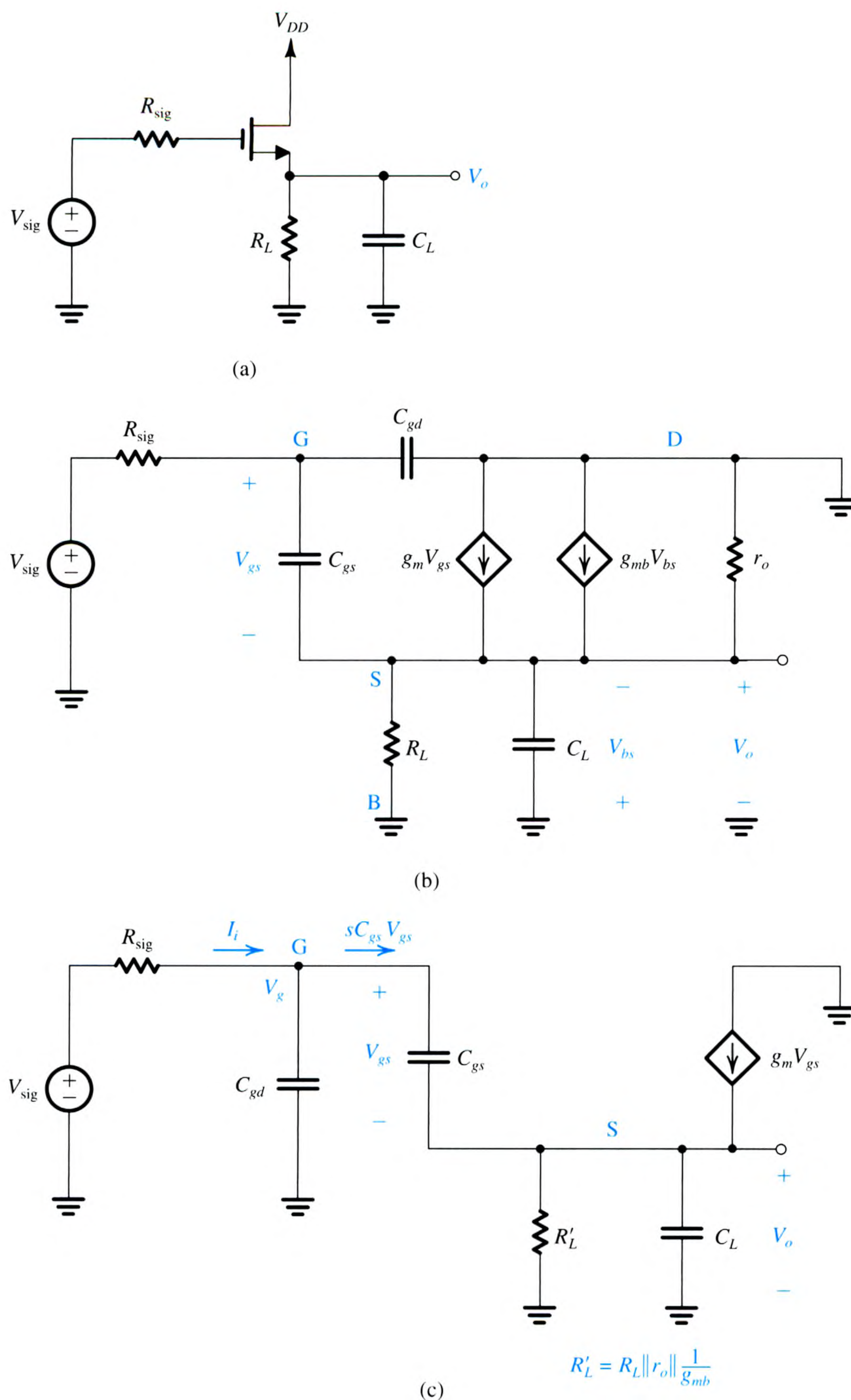


Figure 10.23 (**a**) A directly coupled source follower without the bias detail; (**b**) high-frequency equivalent circuit of the source follower; (**c**) a simplified version of the equivalent circuit.

3. A node equation at S gives

$$(g_m + sC_{gs})V_{gs} = \left(\frac{1}{R_L'} + sC_L\right)V_o$$

which can be used to express V_{gs} in terms of V_o as

$$V_{gs} = \frac{1}{g_m R_L'} \frac{1 + sC_L R_L'}{1 + s(C_{gs}/g_m)} V_o \tag{10.99}$$

Substituting this expression of V_{gs} into Eq. (10.98) results in an equation containing only V_o and V_{sig}; this equation can be manipulated to obtain the source-follower transfer function in the form

$$\frac{V_o}{V_{sig}}(s) = A_M \frac{1 + \left(\frac{s}{\omega_Z}\right)}{1 + b_1 s + b_2 s^2} \tag{10.100}$$

where

$$A_M = \frac{R_L'}{R_L' + \frac{1}{g_m}} = \frac{g_m R_L'}{g_m R_L' + 1} \tag{10.101}$$

$$\omega_Z = g_m / C_{gs} \tag{10.102}$$

$$b_1 = \left(C_{gd} + \frac{C_{gs}}{g_m R_L' + 1}\right)R_{sig} + \left(\frac{C_{gs} + C_L}{g_m R_L' + 1}\right)R_L' \tag{10.103}$$

$$b_2 = \frac{(C_{gs} + C_{gd})C_L + C_{gs}C_{gd}}{g_m R_L' + 1} R_{sig} R_L' \tag{10.104}$$

Analysis of the Source-Follower Transfer Function We now make a number of observations on the transfer function in Eq. (10.100), which describes the gain of the source follower at high frequencies:

1. Since the source follower in Fig. 10.23(a) is directly coupled, the gain at dc is equal to A_M. This correlates with Eq. (10.100) as

$$A_M = V_o/V_{sig}|_{s=0}$$

2. Although the equivalent circuit of Fig. 10.23(c) has three capacitors, the transfer function is of second order because the three capacitors form a continuous loop.
3. The two transmission zeros can be found from Eq. (10.100) as the values of s for which $V_o/V_{sig} = 0$. From Eq. (10.100), we see that V_o/V_{sig} approaches 0 as s approaches ∞. Thus one transmission zero is at $s = \infty$. Physically, this zero is a result of C_{gd}, which appears across the input terminals, becoming a short circuit at infinite frequency and thus making $V_o = 0$. From the numerator of Eq. (10.100) we see that the other transmission zero is at $s = -\omega_Z$ where ω_Z is given by Eq. (10.102). We note that ω_Z is slightly higher than the unity-gain frequency ω_T of the MOSFET [Eq. (10.9)],

$$\omega_T = \frac{g_m}{C_{gs} + C_{gd}} \tag{10.105}$$

Thus the finite transmission zero is at such a high frequency that its effect on the frequency response of the follower is negligibly small.

4. The two poles of the source follower can be found as the roots of the denominator polynomial $(1+b_1 s+b_2 s^2)$. If the poles are real, their frequencies, say ω_{P1} and ω_{P2}, can be found from

$$1+b_1 s+b_2 s^2 = \left(1+\frac{s}{\omega_{P1}}\right)\left(1+\frac{s}{\omega_{P2}}\right) \tag{10.106}$$

Now if $\omega_{P2} \gg \omega_{P1}$ (at least four times larger), a dominant pole exists with frequency ω_{P1} and the 3-dB frequency f_H is given by

$$f_H \simeq f_{P1} \simeq \frac{1}{2\pi b_1} \tag{10.107}$$

Here we remind the reader that b_1 is also τ_H, the effective high-frequency time constant evaluated in the method of open-circuit time constants.[4]

5. If the poles are real but none is dominant, the 3-dB frequency can be determined analytically from the transfer function as the frequency at which $|V_o/V_{sig}| = A_M/\sqrt{2}$. An approximate value can be obtained using the formula[5]

$$f_H \simeq 1\Bigg/\sqrt{\frac{1}{f_{P1}^2}+\frac{1}{f_{P2}^2}-\frac{2}{f_Z^2}} \tag{10.108}$$

6. If the poles are complex, they are best described in terms of their frequency ω_0 and Q-factor, where

$$1+b_1 s+b_2 s^2 = 1+\frac{1}{Q}\frac{s}{\omega_0}+\frac{s^2}{\omega_0^2} \tag{10.109}$$

and for complex poles, $Q > 0.5$. Figure 10.24(a) provides a geometrical interpretation of ω_0 and Q. Figure 10.24(b) shows a number of possible responses obtained for various values of Q. As indicated, the response of the source follower shows no peaking for $Q \le 0.707$. The boundary case corresponding to $Q = 0.707$ (poles at 45° angles) results in what is known as a **maximally flat response** for which $f_{3dB} = f_0$. In terms of the component values of the source follower, ω_0 and Q are given by

$$\omega_0 = \frac{1}{\sqrt{b_2}} = \sqrt{\frac{g_m R_L' + 1}{R_{sig} R_L'[(C_{gs}+C_{gd})C_L + C_{gs}C_{gd}]}} \tag{10.110}$$

$$Q = \frac{\sqrt{b_2}}{b_1} = \frac{\sqrt{g_m R_L' + 1}\sqrt{[(C_{gs}+C_{gd})C_L + C_{gs}C_{gd}]R_{sig}R_L'}}{[C_{gs}+C_{gd}(g_m R_L' + 1)]R_{sig} + (C_{gs}+C_L)R_L'} \tag{10.111}$$

[4]We encourage you to find the three open-circuit time constants, τ_{gd}, τ_{gs}, and τ_{CL} from the circuit in Fig. 10.23(c), and show that their sum τ_H is equal to b_1 in Eq. (10.103).

[5]Derivation of this formula is straightforward and can be found in earlier editions of this book.

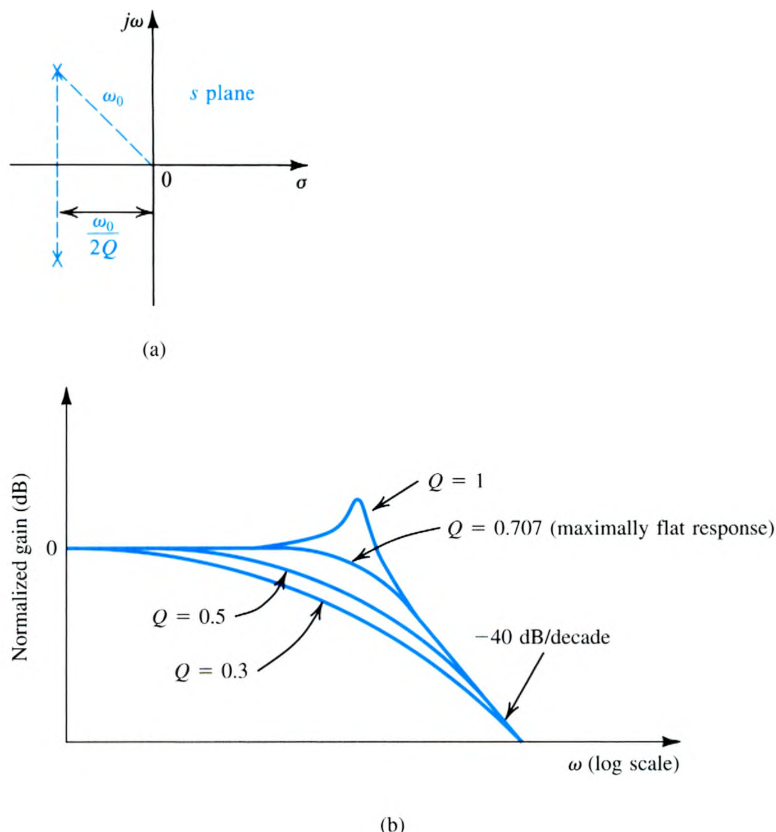


Figure 10.24 (a) A pair of complex-conjugate poles with the definition of ω_0 and Q indicated. (b) Magnitude response of a source (or emitter) follower for different values of the parameter Q. Note that the response is normalized relative to A_M.

Example 10.8

A source follower operated at $g_m = 2$ mA/V and $r_o = 20\,\text{k}\Omega$ is fed with a signal source for which $R_{\text{sig}} = 10\,\text{k}\Omega$ and is loaded in a resistance $R_L = 20\,\text{k}\Omega$. The MOSFET has $C_{gs} = 20$ fF, $C_{gd} = 5$ fF, and $g_{mb} = \chi g_m$ where $\chi = 0.2$, and the total capacitance at the output $C_L = 15$ fF. Determine $A_M, f_T, f_Z, Q, f_{P1}, f_{P2}$, and $f_{3\text{dB}}$.

Solution

$$R'_L = R_L \parallel r_o \parallel \frac{1}{g_{mb}}$$

$$= 20 \parallel 20 \parallel \frac{1}{0.2 \times 2} = 20 \parallel 20 \parallel 2.5 = 2\,\text{k}\Omega$$

$$A_M = \frac{R'_L}{R'_L + \dfrac{1}{g_m}} = \frac{2}{2 + \dfrac{1}{2}} = 0.8 \text{ V/V}$$

$$f_T = \frac{g_m}{2\pi(C_{gs}+C_{gd})}$$
$$= \frac{2\times 10^{-3}}{2\pi(20+5)\times 10^{-15}}$$
$$= 12.7\text{ GHz}$$

$$f_Z = \frac{g_m}{2\pi C_{gs}} = \frac{2\times 10^{-3}}{2\pi\times 20\times 10^{-15}} = 15.9\text{ GHz}$$

To evaluate Q we substitute the given component values into Eq. (10.111) to get

$$Q = 0.42$$

Thus the poles are real. Their frequencies can be obtained by finding the roots of the polynomial $(1+b_1s+b_2s^2)$, where from Eqs. (10.103) and (10.104),

$$b_1 = \tau_H = 104\text{ps}$$

and

$$b_2 = 1.9\times 10^{-21}$$

Thus,

$$f_{P1} = 1.98\text{ GHz}$$
$$f_{P2} = 6.73\text{ GHz}$$

Since $f_{P2}/f_{P1} = 3.4 < 4$, no dominant pole exists. An approximate value for f_H can be obtained as

$$f_H = 1\Bigg/\sqrt{\frac{1}{f_{P1}^2}+\frac{1}{f_{P2}^2}-\frac{2}{f_Z^2}} = 1.93\text{GHz}$$

The exact value of f_H can be found from the transfer function as 1.86 GHz, which is not much different from the approximate value.

EXERCISES

10.18 Recalling that $\tau_H = b_1$, use the expression for b_1 in Eq. (10.103) to find expressions for the three resistances R_{gs}, R_{gd}, and R_{C_L} for the source follower.

Ans. $R_{gs} = \dfrac{R_{sig}+R_L'}{g_mR_L'+1}$; $R_{gd} = R_{sig}$; $R_{C_L} = \dfrac{R_L'}{g_mR_L'+1}$

10.19 In Example 10.8, even though we found that a dominant pole does not exist, use the method of open-circuit time constants to obtain an estimate for f_H. (*Hint*: Recall that $\tau_H = b_1$.)

Ans. $f_H = 1.53$ GHz; about 18% lower than the exact value of 1.86 GHz; still not a bad estimate!

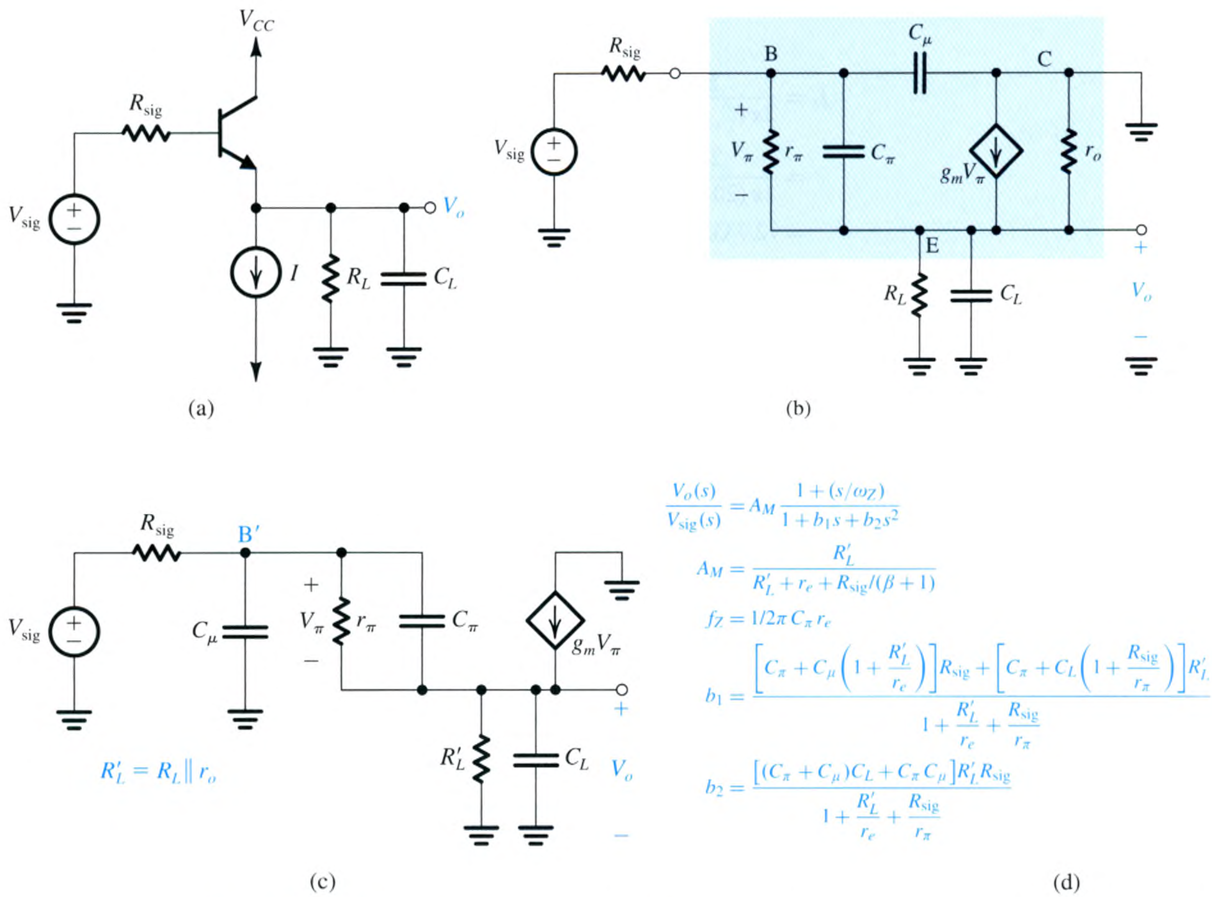


Figure 10.25 **(a)** Emitter follower. **(b)** High-frequency equivalent circuit. **(c)** Simplified equivalent circuit. **(d)** Transfer function.

10.5.2 The Emitter-Follower Case

Figure 10.25 provides the results for the case of the emitter follower. The analysis here is a little more complicated because of the finite β of the BJT.

EXERCISE

10.20 For an emitter follower biased at $I_C = 1$ mA and having $R_{sig} = R_L = 1$ kΩ, $r_o = 100$ kΩ, $\beta = 100$, $C_\mu = 2$ pF, $C_L = 0$, and $f_T = 400$ MHz, find the low-frequency gain A_M, f_Z, f_{P1}, f_{P2}, and an estimate for f_H.

Ans. 0.97 V/V; 458 MHz; 67.2 MHz; 562 MHz; 67.2 MHz

A

10.6 High-Frequency Response of Differential Amplifiers

In this section we will consider the variation with frequency of both the differential gain and the common-mode gain and hence of the CMRR. We will rely heavily on the study of frequency response of single-ended amplifiers presented in the sections above, but we will consider MOS circuits only; the bipolar case is a straightforward extension, as we saw above on a number of occasions.

10.6.1 Analysis of the Resistively Loaded MOS Amplifier

We begin with the basic, resistively loaded MOS differential pair shown in Fig. 10.26(a). Note that we have explicitly shown transistor Q_S that supplies the bias current I. Although we are showing a dc bias voltage V_{BIAS} at its gate, usually Q_S is part of a current mirror. This detail, however, is of no consequence to our present needs. Most importantly, we are interested in the total impedance between node S and ground, Z_{SS}, because this impedance plays a significant role in determining the common-mode gain and the CMRR of the differential amplifier. Resistance R_{SS} is simply the output resistance of current source Q_S. Capacitance C_{SS} is the total capacitance between node S and ground and includes C_{db} and C_{gd} of Q_S, as well as C_{sb1} and C_{sb2}. This capacitance can be significant, especially if wide transistors are used for Q_S, Q_1, and Q_2.

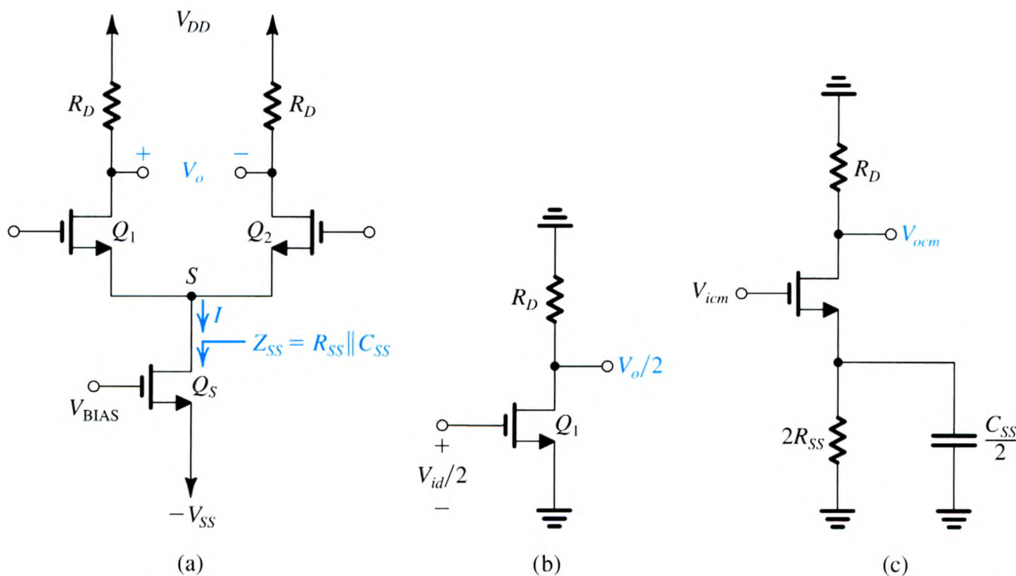


Figure 10.26 **(a)** A resistively loaded MOS differential pair; the transistor supplying the bias current is explicitly shown. It is assumed that the total impedance between node S and ground, Z_{SS}, consists of a resistance R_{SS} in parallel with a capacitance C_{SS}. **(b)** Differential half-circuit. **(c)** Common-mode half-circuit.

The differential half-circuit shown in Fig. 10.26(b) can be used to determine the frequency dependence of the differential gain V_o/V_{id}. Indeed the gain function $A_d(s)$ of the differential amplifier will be identical to the transfer function of this common-source amplifier.[6] We studied the frequency response of the common-source amplifier at great length in Sections 10.2 and 10.3 and will not repeat this material here.

EXERCISE

10.21 A MOSFET differential amplifier such as that in Fig. 10.26(a) is biased with a current $I = 0.8$ mA. The transistors Q_1 and Q_2 have $W/L = 100$, $k_n' = 0.2$ mA/V^2, $V_A = 20$ V, $C_{gs} = 50$ fF, $C_{gd} = 10$ fF, and $C_{db} = 10$ fF. The drain resistors are 5 kΩ each. Also, there is a 100-fF capacitive load between each drain and ground.

(a) Find V_{OV} and g_m for each transistor.

(b) Find the differential gain A_d.

(c) If the input signal source has a small resistance R_{sig} and thus the frequency response is determined primarily by the output pole, estimate the 3-dB frequency f_H. [*Hint:* Refer to Section 10.2.3 and specifically to Eq. (10.45).]

(d) If, in a different situation, the amplifier is fed symmetrically with a signal source of 20 kΩ resistance (i.e., 10 kΩ in series with each gate terminal), use the open-circuit time-constants method to estimate f_H. [*Hint:* Refer to Section 10.3.3 and specifically to Eq. (10.68).]

Ans. (a) 0.2 V, 4 mA/V; (b) 18.2 V/V; (c) 292 MHz; (d) 53.7 MHz

The common-mode half-circuit is shown in Fig. 10.26(c). Although this circuit has other capacitances, namely, C_{gs}, C_{gd}, and C_{db} of the transistor in addition to other stray capacitances, we have chosen to show only $C_{SS}/2$. This is because $(C_{SS}/2)$ together with $(2R_{SS})$ form a real-axis zero in the common-mode gain function at a frequency much lower than those of the other poles and zeros of the circuit. This zero then dominates the frequency dependence of A_{cm} and CMRR.

If the output of the differential amplifier is taken single-endedly, then the common-mode gain of interest is V_{ocm}/V_{icm}. More typically, the output is taken differentially. Nevertheless, as we have seen in Section 9.3, V_{ocm}/V_{icm} still plays a major role in determining the common-mode gain. To be specific, consider what happens when the output is taken differentially and there is a mismatch ΔR_D between the two drain resistances. The resulting common-mode gain was found in Section 9.3 to be (Eq. 9.80)

$$A_{cm} = -\left(\frac{R_D}{2R_{SS}}\right)\frac{\Delta R_D}{R_D} \tag{10.112}$$

[6]Here we are not showing the resistance of the signal source R_{sig}, which, of course, must be included in the frequency-response analysis, as we have done in the case of the CS amplifier. See Exercise 10.21.

which is simply the product of V_{ocm}/V_{icm} and the per-unit mismatch $(\Delta R_D/R_D)$. Similar expressions can be found for the effects of other circuit mismatches. The important point to note is that the factor $R_D/2R_{SS}$ is always present in these expressions. Thus, the frequency dependence of A_{cm} can be obtained by simply replacing R_{SS} by Z_{SS} in this factor. Doing so for the expression in Eq. (10.112) gives

$$\begin{aligned} A_{cm}(s) &= -\frac{R_D}{2Z_{SS}}\left(\frac{\Delta R_D}{R_D}\right) \\ &= -\frac{1}{2}R_D\left(\frac{\Delta R_D}{R_D}\right)Y_{SS} \\ &= -\frac{1}{2}R_D\left(\frac{\Delta R_D}{R_D}\right)\left(\frac{1}{R_{SS}}+sC_{SS}\right) \\ &= -\frac{R_D}{2R_{SS}}\left(\frac{\Delta R_D}{R_D}\right)(1+sC_{SS}R_{SS}) \end{aligned} \tag{10.113}$$

from which we see that A_{cm} acquires a zero on the negative real axis of the s plane with frequency ω_Z,

$$\omega_Z = \frac{1}{C_{SS}R_{SS}} \tag{10.114}$$

or in hertz,

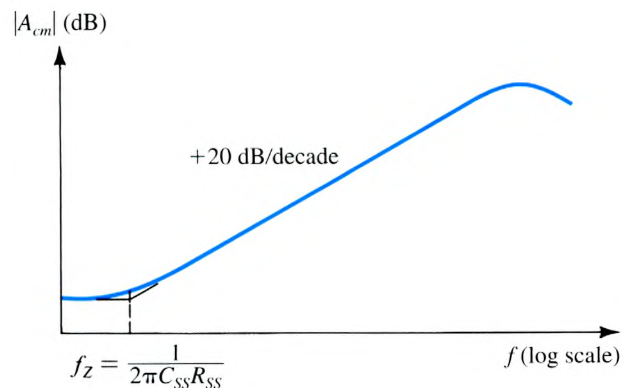
$$f_Z = \frac{\omega_Z}{2\pi} = \frac{1}{2\pi C_{SS}R_{SS}} \tag{10.115}$$

As mentioned above, usually f_Z is much lower than the frequencies of the other poles and zeros. As a result, the common-mode gain increases at the rate of +20 dB/decade starting at a relatively low frequency, as shown in Fig. 10.27(a). Of course, A_{cm} drops off at high frequencies because of the other poles of the common-mode half-circuit. It is, however, f_Z that is significant, for it is the frequency at which the CMRR of the differential amplifier begins to decrease, as indicated in Fig. 10.27(c). Note that if both A_d and A_{cm} are expressed and plotted in dB, then CMRR in dB is simply the difference between A_d and A_{cm}.

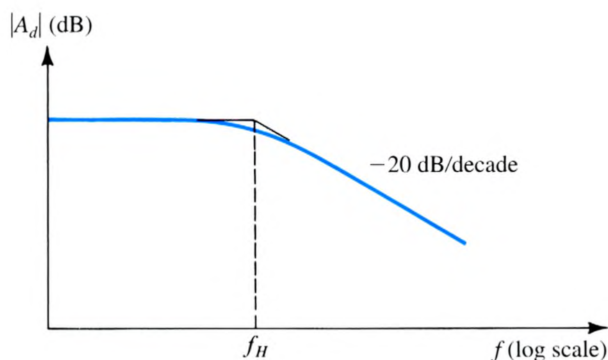
Although in the foregoing we considered only the common-mode gain resulting from an R_D mismatch, the results apply to the common-mode gain resulting from any other mismatch. For instance, it applies equally well to the case of a g_m mismatch, modifying Eq. (9.84) by replacing R_{SS} by Z_{SS}, and so on.

Before leaving this section, it is interesting to point out an important trade-off found in the design of the current-source transistor Q_S: In order to operate this current source with a small V_{DS} (to conserve the already low V_{DD}), we desire to operate the transistor at a low overdrive voltage V_{OV}. For a given value of the current I, however, this means using a large W/L ratio (i.e., a wide transistor). This in turn increases C_{SS} and hence lowers f_Z with the result that the CMRR deteriorates (i.e., decreases) at a relatively low frequency. Thus there is a trade-off between the need to reduce the dc voltage across Q_S and the need to keep the CMRR reasonably high at higher frequencies.

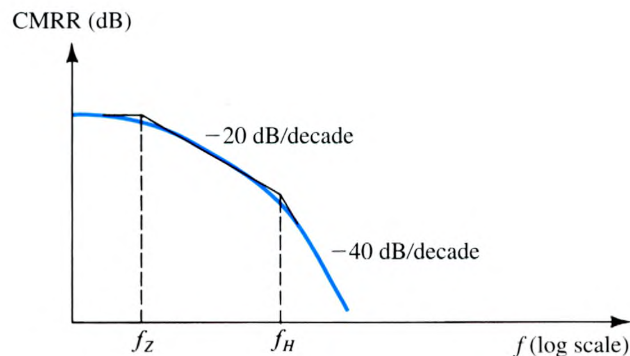
To appreciate the need for high CMRR at higher frequencies, consider the situation illustrated in Fig. 10.28: We show two stages of a differential amplifier whose power-supply voltage V_{DD} is corrupted with high-frequency noise. Since the quiescent voltage at each



(a)



(b)



(c)

Figure 10.27 Variation of **(a)** common-mode gain, **(b)** differential gain, and **(c)** common-mode rejection ratio with frequency.

of the drains of Q_1 and Q_2 is $[V_{DD} - (I/2)R_D]$ we see that v_{D1} and v_{D2} will have the same high-frequency noise as V_{DD}. This high-frequency noise then constitutes a common-mode input signal to the second differential stage, formed by Q_3 and Q_4. If the second differential stage is perfectly matched, its differential output voltage V_o should be free of high-frequency noise. However, in practice there is no such thing as perfect matching, and the second stage

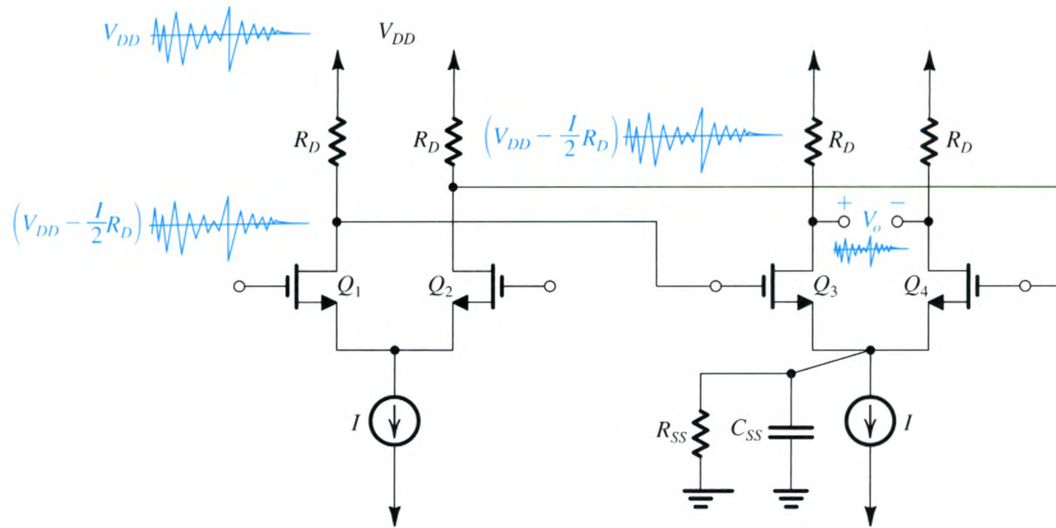


Figure 10.28 The second stage in a differential amplifier, which is relied on to suppress high-frequency noise injected by the power supply of the first stage, and therefore must maintain a high CMRR at higher frequencies.

will have a finite common-mode gain. Furthermore, because of the zero formed by R_{SS} and C_{SS} of the second stage, the common-mode gain will increase with frequency, causing some of the noise to make its way to V_o. With careful design, this undesirable component of V_o can be kept small.

EXERCISE

10.22 The differential amplifier specified in Exercise 10.21 has $R_{SS} = 75\text{ k}\Omega$ and $C_{SS} = 0.4\text{ pF}$. Find the 3-dB frequency of the CMRR.
Ans. 5.3 MHz

RFID: IDENTIFICATION AT A DISTANCE

Radio-frequency identification (RFID) tags are increasingly important elements in our daily lives. These tiny devices embody a wide range of electronic techniques: linear, digital, radio, power, and signaling, all at very low power levels. They are activated (and powered) by an external interrogation signal that can be located a few feet away. Once activated, they respond with the requested data, most often simply an identifying number. Applications are incredibly diverse, ranging from automobile key fobs to tags that permit the recovery of lost pets.

10.6.2 Frequency Response of the Current-Mirror-Loaded MOS Differential Amplifier

We will now consider the frequency response of the current-mirror-loaded MOS differential-pair circuit studied in Section 9.5. The circuit is shown in Fig. 10.29(a) with two capacitances indicated: C_m, which is the total capacitance at the input node of the current mirror, and C_L, which is the total capacitance at the output node of the amplifier. Capacitance C_m is mainly formed by C_{gs3} and C_{gs4} but also includes C_{gd1}, C_{db1}, and C_{db3},

$$C_m = C_{gd1} + C_{db1} + C_{db3} + C_{gs3} + C_{gs4} \tag{10.116}$$

Capacitance C_L includes C_{gd2}, C_{db2}, C_{db4}, and C_{gd4} as well as an actual load capacitance and/or the input capacitance of a subsequent stage (C_x),

$$C_L = C_{gd2} + C_{db2} + C_{gd4} + C_{db4} + C_x \tag{10.117}$$

These two capacitances primarily determine the dependence of the differential gain of this amplifier on frequency.

In the following we show that C_L introduces the dominant pole of the differential gain. We will also show that capacitance C_m affects the overall short-circuit transconductance G_m and hence the differential gain, introducing a pole and a zero at much higher frequencies.

The Dominant Pole Figure 10.29(b) shows the output equivalent circuit of the differential amplifier, from which we can write

$$\frac{V_o}{V_{id}} = \frac{G_m R_o}{1 + sC_L R_o} \tag{10.118}$$

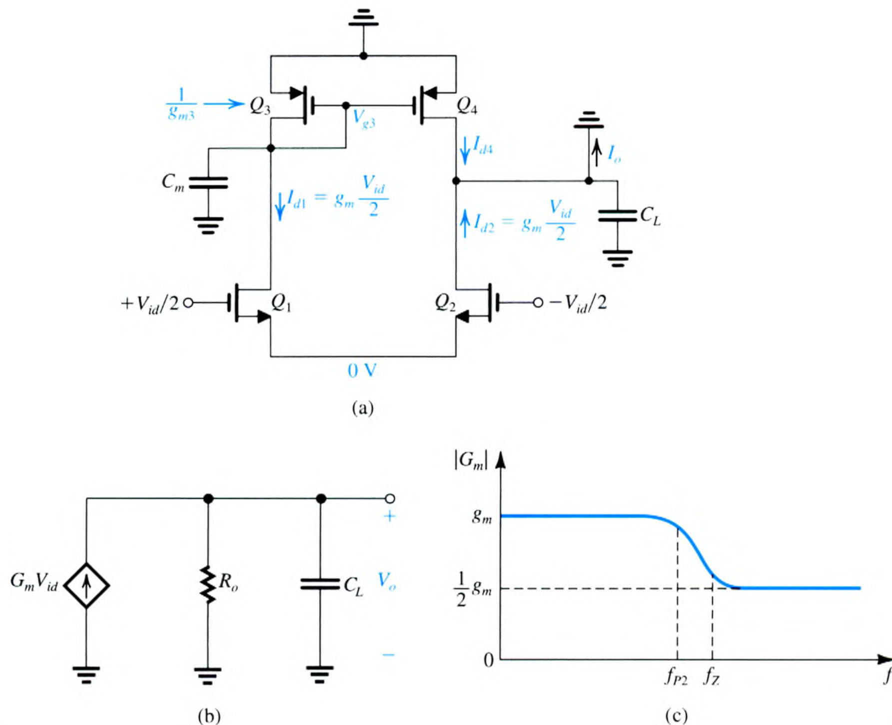


Figure 10.29 **(a)** Frequency–response analysis of the active-loaded MOS differential amplifier. **(b)** Output equivalent circuit. **(c)** The overall transconductance G_m as a function of frequency.

Neglecting for the time being the frequency dependence of G_m, then

$$G_m = g_{m1} = g_{m2}$$

and

$$R_o = r_{o2} \| r_{o4}$$

Equation (10.118) indicates that the differential gain has a pole at

$$f_{P1} = \frac{1}{2\pi C_L R_o} \tag{10.119}$$

This pole is usually at a much lower frequency than the pole and zero introduced by C_m and thus is dominant.

Dependence of G_m on Frequency To see the dependence of G_m on frequency, refer to Fig. 10.29(a), which shows the circuit with a short circuit at the output terminal. The short-circuit transconductance G_m can be found as

$$G_m = \frac{I_o}{V_{id}} \tag{10.120}$$

Now, I_o is the sum of I_{d2} and I_{d4},

$$I_o = I_{d2} + I_{d4} \tag{10.121}$$

Here, while I_{d2} is simply $g_m(V_{id}/2)$, current I_{d4} is obtained through the action of the current mirror and is thus affected by capacitance C_m. Specifically,

$$I_{d4} = -g_{m4} V_{g3} \tag{10.122}$$

where V_{g3} can be found from

$$V_{g3} = -I_{d1} \frac{1}{\dfrac{1}{R_m} + sC_m} \tag{10.123}$$

where R_m is the total resistance at the input node of the current mirror,

$$R_m = r_{o1} \| \frac{1}{g_{m3}} \| r_{o3} \simeq 1/g_{m3} \tag{10.124}$$

Substituting this approximate value of R_m into Eq. (10.123) and substituting $I_{d1} = g_m(V_{id}/2)$ gives

$$V_{g3} = -\frac{g_m(V_{id}/2)}{g_{m3} + sC_m}$$

Substituting this expression for V_{g3} into Eq. (10.122) gives

$$I_{d4} = \frac{g_{m4} g_m (V_{id}/2)}{g_{m3} + sC_m}$$

Since $g_{m3} = g_{m4}$, this equation reduces to

$$I_{d4} = \frac{g_m V_{id}/2}{1 + s\frac{C_m}{g_{m3}}} \tag{10.125}$$

Substituting for I_{d4} from Eq. (10.125) and $I_{d2} = g_m(V_{id}/2)$ into Eq. (10.121) gives

$$I_o = g_m(V_{id}/2) + \frac{g_m V_{id}/2}{1 + s\frac{C_m}{g_{m3}}} \tag{10.126}$$

We can now obtain G_m as

$$G_m \equiv \frac{I_o}{V_{id}} = g_m \frac{1 + s\dfrac{C_m}{2g_{m3}}}{1 + s\dfrac{C_m}{g_{m3}}} \tag{10.127}$$

Thus, as expected, the low-frequency value of G_m is equal to g_m of Q_1 and Q_2. At high frequencies, G_m acquires a pole and a zero, the frequencies of which are

$$f_{P2} = \frac{g_{m3}}{2\pi C_m} \tag{10.128}$$

and

$$f_Z = \frac{2g_{m3}}{2\pi C_m} \tag{10.129}$$

That is, the zero frequency is twice that of the pole. Since C_m is approximately equal to $C_{gs2} + C_{gs4} = 2C_{gs}$, we also have

$$f_{P2} = \frac{g_{m3}}{2\pi C_m} \simeq \frac{g_{m3}}{2\pi (2C_{gs})} \simeq f_T/2 \tag{10.130}$$

and

$$f_Z \simeq f_T \tag{10.131}$$

where f_T is the transition frequency of the MOSFET Q_3. Thus, the **mirror pole and zero** occur at very high frequencies.

Figure 10.29(c) shows a sketch of the magnitude of G_m versus frequency. Notice that at very high frequencies, C_m acts as a short circuit, reducing V_{g3} to zero. As a result, I_{d4} becomes zero, and I_0 becomes equal to I_{d1} only, which is half of the low-frequency value of I_0. The result is that G_m is reduced to half of its value, as shown in the sketch in Fig. 10.29(c). The complete gain function can be expressed as

$$\frac{V_o}{V_{id}} = (g_m R_o)\left(\frac{1 + s\dfrac{C_m}{2g_{m3}}}{1 + s\dfrac{C_m}{g_{m3}}}\right)\left(\frac{1}{1 + sC_L R_o}\right) \tag{10.132}$$

Example 10.9

Consider a current-mirror-loaded MOS differential amplifier of the type shown in Fig. 10.29(a). Assume that for all transistors, $W/L = 7.2\ \mu\text{m}/0.36\ \mu\text{m}$, $C_{gs} = 20$ fF, $C_{gd} = 5$ fF, and $C_{db} = 5$ fF. Also, let $\mu_n C_{ox} = 387\ \mu\text{A/V}^2$, $\mu_p C_{ox} = 86\ \mu\text{A/V}^2$, $V'_{An} = 5$ V/μm, and $|V'_{Ap}| = 6$ V/μm. The bias current $I = 0.2$ mA, and the bias current source has an output resistance $R_{SS} = 25$ kΩ and an output capacitance $C_{SS} = 0.2$ pF. In addition to the capacitances introduced by the transistors at the output node, there is a capacitance C_x

of 25 fF. Determine the low-frequency values of A_d, A_{cm}, and CMRR. Also, find the poles and zero of A_d and the dominant pole of CMRR.

Solution

Since $I = 0.2$ mA, each of the four transistors is operating at a bias current of 100 μA. Thus, for Q_1 and Q_2,

$$100 = \frac{1}{2} \times 387 \times \frac{7.2}{0.36} \times V_{OV}^2$$

which leads to

$$V_{OV} = 0.16 \text{ V}$$

Thus,

$$g_m = g_{m1} = g_{m2} = \frac{2 \times 0.1}{0.16} = 1.25 \text{ mA/V}$$

$$r_{o1} = r_{o2} = \frac{5 \times 0.36}{0.1} = 18 \text{ k}\Omega$$

For Q_3 and Q_4 we have

$$100 = \frac{1}{2} \times 86 \times \frac{7.2}{0.36} V_{OV3,4}^2$$

Thus,

$$|V_{OV3,4}| = 0.34 \text{ V}$$

and

$$g_{m3} = g_{m4} = \frac{2 \times 0.1}{0.34} = 0.6 \text{ mA/V}$$

$$r_{o3} = r_{o4} = \frac{6 \times 0.36}{0.1} = 21.6 \text{ k}\Omega$$

The low-frequency value of the differential gain can be determined from

$$\begin{aligned} A_d &= g_m(r_{o2} \parallel r_{o4}) \\ &= 1.25(18 \parallel 21.6) = 12.3 \text{ V/V} \end{aligned}$$

The low-frequency value of the common-mode gain can be determined from Eq. (9.137) as

$$\begin{aligned} A_{cm} &= -\frac{1}{2g_{m3}R_{SS}} \\ &= -\frac{1}{2 \times 0.6 \times 25} = -0.033 \text{ V/V} \end{aligned}$$

Example 10.9 *continued*

The low-frequency value of the CMRR can now be determined as

$$\text{CMRR} = \frac{|A_d|}{|A_{cm}|} = \frac{12.3}{0.033} = 369$$

or,

$$20\log 369 = 51.3 \text{ dB}$$

To determine the poles and zero of A_d we first compute the values of the two pertinent capacitances C_m and C_L. Using Eq. (10.116),

$$\begin{aligned} C_m &= C_{gd1} + C_{db1} + C_{db3} + C_{gs3} + C_{gs4} \\ &= 5 + 5 + 5 + 20 + 20 = 55 \text{ fF} \end{aligned}$$

Capacitance C_L is found using Eq. (10.117) as

$$\begin{aligned} C_L &= C_{gd2} + C_{db2} + C_{gd4} + C_{db4} + C_x \\ &= 5 + 5 + 5 + 5 + 25 = 45 \text{ fF} \end{aligned}$$

Now, the poles and zero of A_d can be found from Eqs. (10.119), (10.128), and (10.129) as

$$\begin{aligned} f_{P1} &= \frac{1}{2\pi C_L R_o} \\ &= \frac{1}{2\pi \times C_L(r_{o2} \| r_{o4})} \\ &= \frac{1}{2\pi \times 45 \times 10^{-15}(18 \| 21.6)10^3} \\ &= 360 \text{ MHz} \end{aligned}$$

$$f_{P2} = \frac{g_{m3}}{2\pi C_m} = \frac{0.6 \times 10^{-3}}{2\pi \times 55 \times 10^{-15}} = 1.74 \text{ GHz}$$

$$f_Z = 2f_{P2} = 3.5 \text{ GHz}$$

Thus the dominant pole is that produced by C_L at the output node. As expected, the pole and zero of the mirror are at much higher frequencies.

The dominant pole of the CMRR is at the location of the common-mode-gain zero introduced by C_{SS} and R_{SS}, that is,

$$\begin{aligned} f_Z &= \frac{1}{2\pi C_{SS} R_{SS}} \\ &= \frac{1}{2\pi \times 0.2 \times 10^{-12} \times 25 \times 10^3} \\ &= 31.8 \text{ MHz} \end{aligned}$$

Thus, the CMRR begins to decrease at 31.8 MHz, which is much lower than f_{P1}.

EXERCISE

10.23 A bipolar current-mirror-loaded differential amplifier is biased with a current source $I = 1$ mA. The transistors are specified to have $|V_A| = 100$ V. The total capacitance at the output node is 2 pF. Find the dc value, and the frequency of the dominant high-frequency pole, of the differential voltage gain.
Ans. 2000 V/V; 0.8 MHz

10.7 Other Wideband Amplifier Configurations



Thus far, we have studied one wideband amplifier configuration: the cascode amplifier. In Section 10.4, we found that adding the cascode transistor can enable us to reduce the gain required of the common-source transistor and hence reduce the Miller effect and, correspondingly, increase f_H. In this section, we study three other circuit configurations that obtain increased f_H by decreasing or eliminating the Miller effect.

10.7.1 Obtaining Wideband Amplification by Source or Emitter Degeneration

As we discussed in Chapter 7, adding a resistance in the source (emitter) lead of a CS (CE) amplifier can result in a number of performance improvements at the expense of a reduction in voltage gain. Increasing the amplifier bandwidth f_H, which is the topic of interest to us in this section, is among those improvements.

To see how this comes about, consider the source-degenerated common-source amplifier in Fig. 10.30(a) and its output equivalent circuit in Fig. 10.30(b). For the usual case of $g_m r_o \gg 1$, the short-circuit transconductance G_m can be shown to be given by

$$G_m \simeq \frac{g_m}{1 + g_m R_s} \tag{10.133}$$

and the output resistance is given by Eq. (8.60), that is,

$$R_o \simeq r_o(1 + g_m R_s) \tag{10.134}$$

Thus, source degeneration reduces the transconductance and increases the output resistance by the same factor, $(1 + g_m R_s)$. The low-frequency voltage gain can be obtained as

$$A_M = \frac{V_o}{V_{sig}} = -G_m(R_o \| R_L) = -G_m R_L' \tag{10.135}$$

where

$$R_L' = R_L \| R_o \tag{10.136}$$

The magnitude of A_M will be smaller than that obtained without R_s.

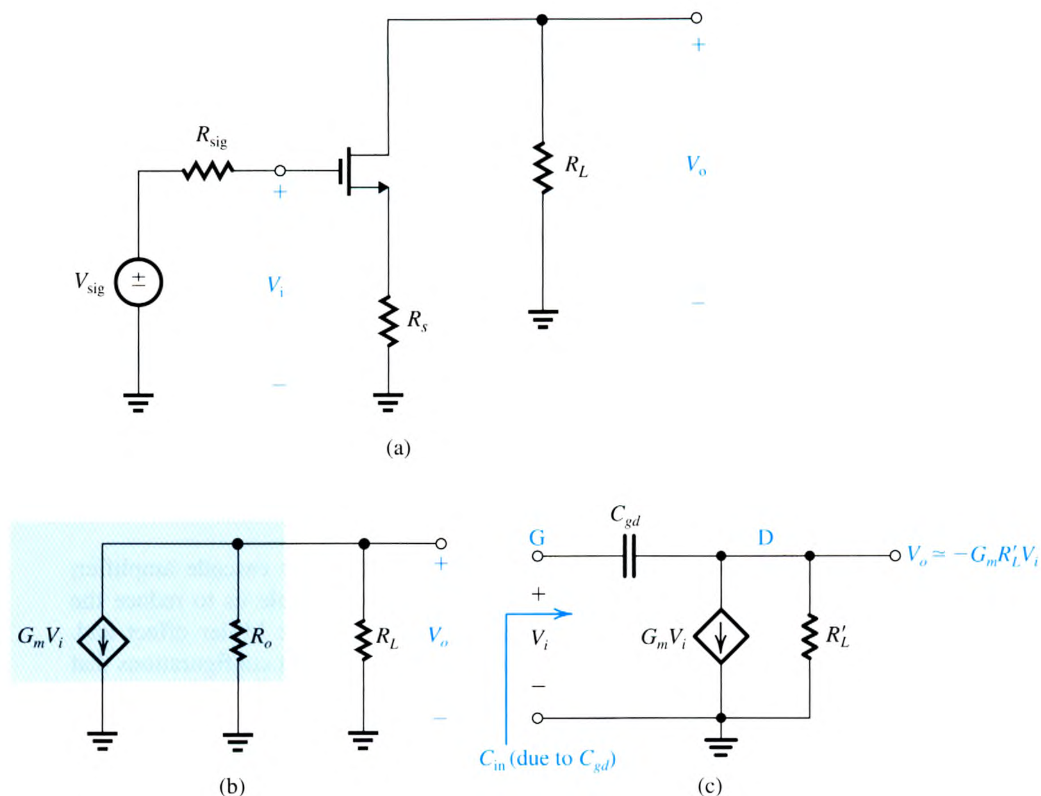


Figure 10.30 **(a)** The CS amplifier circuit, with a source resistance R_s. **(b)** Equivalent-circuit representation of the amplifier output. **(c)** Determining the component of C_{in} due to C_{gd}.

A first cut at finding the effect of R_s on the high-frequency response can be obtained by finding the input capacitance C_{in} due to C_{gd}. Referring to Fig. 10.30(c) and recalling the corresponding analysis of the CS amplifier in Section 10.2.2, we can write

$$C_{in}\ (\text{due to } C_{gd}) = C_{gd}\,(1 + G_mR'_L) \simeq |A_M|C_{gd} \tag{10.137}$$

Since this is usually the dominant component of the input capacitance, and assuming that the 3-dB frequency is determined by the interaction of R_{sig} and C_{in}, we can obtain an approximate estimate of f_H as

$$f_H \simeq \frac{1}{2\pi R_{sig}C_{in}} \simeq \frac{1}{2\pi R_{sig}C_{gd}|A_M|} \tag{10.138}$$

This expression clearly shows that as $|A_M|$ is reduced by including R_s, f_{3dB} is correspondingly increased. Another interesting observation about Eq. (10.137) is that the gain–bandwidth product remains constant as R_s is changed,

$$GB \equiv |A_M|f_H = \frac{1}{2\pi R_{sig}C_{gd}} \tag{10.139}$$

The above analysis, however, is approximate as we have not taken into account the other capacitances present in the circuit. These are shown in Fig. 10.31(a) where C_L includes all

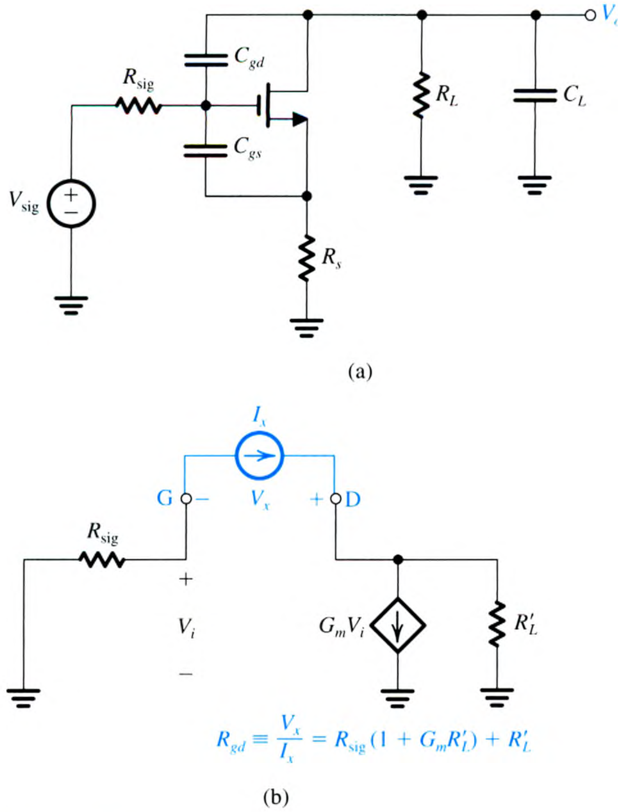


Figure 10.31 **(a)** The source-degenerated CS amplifier with all capacitances shown. **(b)** Determining R_{gd}.

capacitances between the output node and ground. An estimate for f_H can be obtained using the method of open-circuit time constants, as follows:

To obtain R_{gd} we use the circuit in Fig. 10.31(b) and write by analogy to the case without R_s,

$$R_{gd} = R_{sig}(1 + G_m R'_L) + R'_L \tag{10.140}$$

The formula for R_{C_L} can be seen to be simply

$$R_{C_L} = R_L \| R_o = R'_L \tag{10.141}$$

The formula for R_{gs} is the most difficult to derive, and the derivation should be performed with the hybrid-π model explicitly utilized. Straightforward, though somewhat tedious, circuit analysis yields (for $g_m r_o \gg 1$),

$$R_{gs} \simeq \frac{R_{sig} + R_s + R_{sig} R_s / (r_o + R_L)}{1 + g_m R_s \left(\dfrac{r_o}{r_o + R_L} \right)} \tag{10.142}$$

Next we compute τ_H,

$$\tau_H = C_{gs} R_{gs} + C_{gd} R_{gd} + C_L R_{C_L} \tag{10.143}$$

and use it to determine f_H,

$$f_H = \frac{1}{2\pi\,\tau_H}$$

EXERCISE

10.24 Consider a CS amplifier having $g_m = 2$ mA/V, $r_o = 20$ kΩ, $R_L = 20$ kΩ, $R_{sig} = 20$ kΩ, $C_{gs} = 20$ fF, $C_{gd} = 5$ fF, and $C_L = 5$ fF. (a) Find the voltage gain A_M and the 3-dB frequency f_H (using the method of open-circuit time constants) and hence the gain–bandwidth product. (b) Repeat (a) for the case in which a resistance R_s is connected in series with the source terminal with a value selected so that $g_m R_s = 2$.

Ans. (a) –20 V/V, 61.2 MHz, 1.22 GHz; (b) –10 V/V, 109 MHz, 1.1 GHz

10.7.2 Increasing f_H by Buffering the Input Signal Source

Another approach to reducing the effect of the Miller capacitance is to place a buffer (a source follower or an emitter follower) between the input signal source and the input terminal of the common-source or common-emitter amplifier. In this case, while the Miller capacitance in the CS or CE transistor remains unchanged, it will be interacting not with R_{sig} but with the output resistance of the buffer, which is much smaller than R_{sig}. This will result in increasing the frequency of the pole that arises at the amplifier input, and thus extending the bandwidth of the amplifier. Figure 10.32 shows three different implementations of this technique. Note that in addition to increasing f_H, the input buffer raises the input resistance of the common-emitter amplifiers in Figs. 10.32(b) and (c).

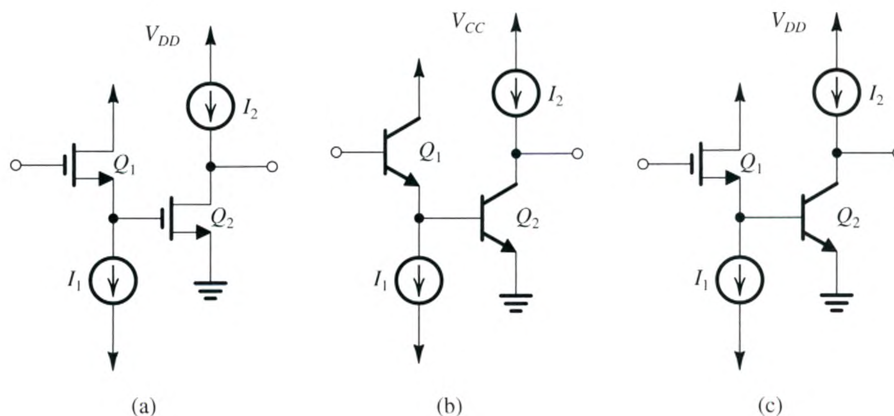


Figure 10.32 **(a)** CD–CS amplifier. **(b)** CC–CE amplifier. **(c)** CD–CE amplifier.

Example 10.10

Consider a CC–CE amplifier such as that in Fig. 10.32(b) with the following specifications: $I_1 = I_2 =$ 1 mA and identical transistors with $\beta = 100, f_T = 400$ MHz, and $C_\mu = 2$ pF. Let the amplifier be fed with a source V_{sig} having a resistance $R_{sig} = 4$ kΩ, and assume a load resistance of 4 kΩ. Find the voltage gain A_M, and estimate the 3-dB frequency, f_H. Compare the results with those obtained with a CE amplifier operating under the same conditions. For simplicity, neglect r_o.

Solution

At an emitter bias current of 1 mA, Q_1 and Q_2 have

$$g_m = 40 \text{ mA/V} \qquad r_e = 25\ \Omega \qquad r_\pi = \frac{\beta}{g_m} = \frac{100}{40} = 2.5 \text{ k}\Omega$$

$$C_\pi + C_\mu = \frac{g_m}{\omega_T} = \frac{g_m}{2\pi f_T} = \frac{40 \times 10^{-3}}{2\pi \times 400 \times 10^6} = 15.9 \text{ pF}$$

$$C_\mu = 2 \text{ pF} \qquad C_\pi = 13.9 \text{ pF}$$

The voltage gain A_M can be determined from the circuit shown in Fig. 10.33(a) as follows:

$$R_{in2} = r_{\pi 2} = 2.5 \text{ k}\Omega$$

$$R_{in} = (\beta_1 + 1)(r_{e1} + R_{in2})$$

$$= 101(0.025 + 2.5) = 255 \text{ k}\Omega$$

$$\frac{V_{b1}}{V_{sig}} = \frac{R_{in}}{R_{in} + R_{sig}} = \frac{255}{255 + 4} = 0.98 \text{ V/V}$$

$$\frac{V_{b2}}{V_{b1}} = \frac{R_{in2}}{R_{in2} + r_{e1}} = \frac{2.5}{2.5 + 0.025} = 0.99 \text{ V/V}$$

$$\frac{V_o}{V_{b2}} = -g_{m2} R_L = -40 \times 4 = -160 \text{ V/V}$$

Thus,

$$A_M = \frac{V_o}{V_{sig}} = -160 \times 0.99 \times 0.98 = -155 \text{ V/V}$$

To determine f_H we use the method of open-circuit time constants. Figure 10.33(b) shows the circuit with V_{sig} set to zero and the four capacitances indicated. Capacitance $C_{\mu 1}$ sees a resistance $R_{\mu 1}$,

$$R_{\mu 1} = R_{sig} \parallel R_{in}$$

$$= 4 \parallel 255 = 3.94 \text{ k}\Omega$$

Example 10.10 *continued*

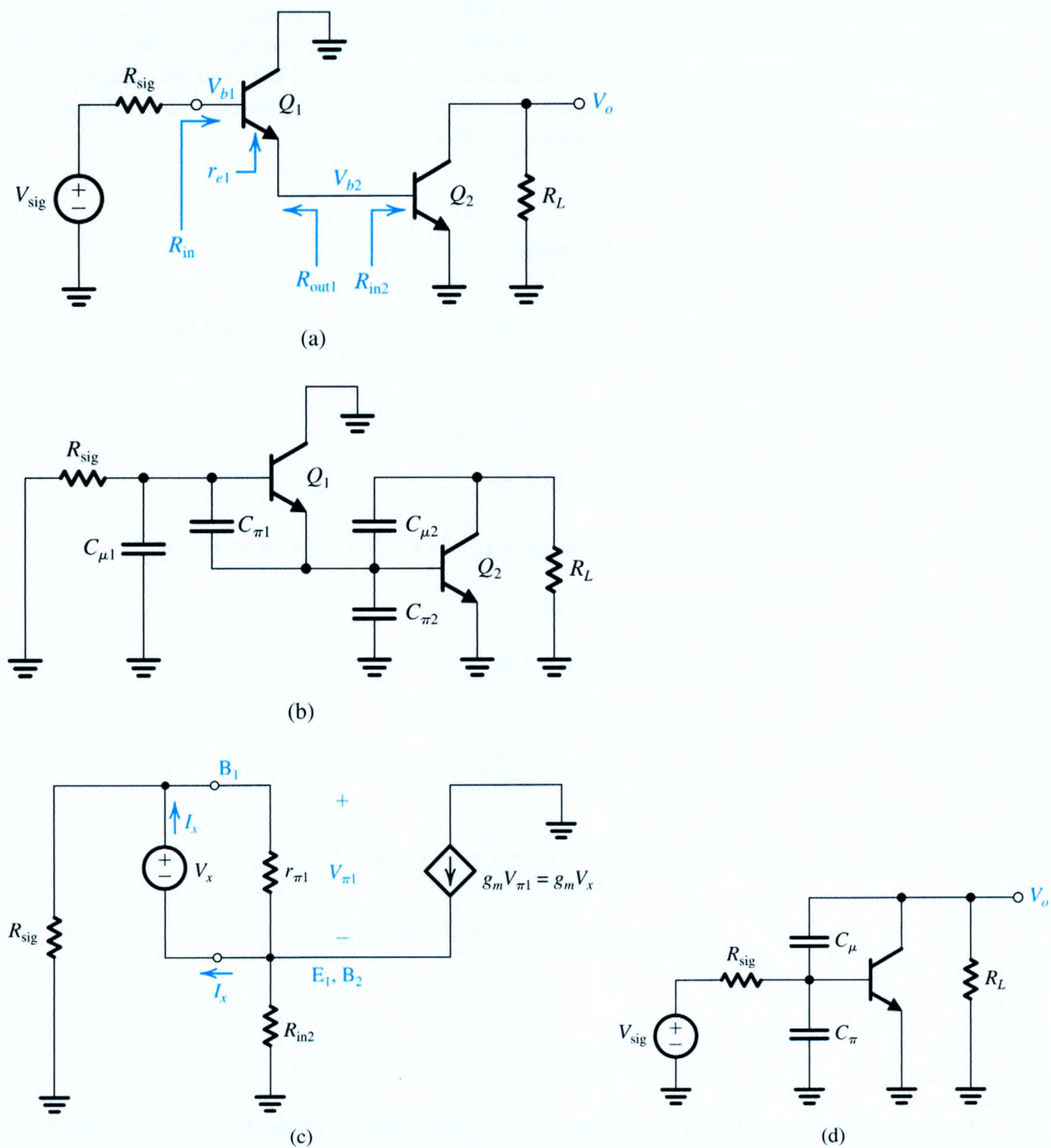


Figure 10.33 Circuits for Example 10.10: **(a)** the CC–CE circuit prepared for low-frequency, small-signal analysis; **(b)** the circuit at high frequencies, with V_{sig} set to zero to enable determination of the open-circuit time constants; **(c)** equivalent circuit for the determination of $R_{\pi 1}$; **(d)** a CE amplifier for comparison.

To find the resistance $R_{\pi 1}$ seen by capacitance $C_{\pi 1}$ we refer to the equivalent circuit in Fig. 10.33(c). Analysis of this circuit results in

$$R_{\pi 1} \equiv \frac{V_x}{I_x} = \frac{R_{sig} + R_{in2}}{1 + \dfrac{R_{sig}}{r_{\pi 1}} + \dfrac{R_{in2}}{r_{e1}}}$$

$$= \frac{4000 + 2500}{1 + \dfrac{4000}{2500} + \dfrac{2500}{25}} = 63.4\ \Omega$$

Capacitance $C_{\pi 2}$ sees a resistance $R_{\pi 2}$,

$$R_{\pi 2} = R_{in2} \parallel R_{out1}$$

$$= r_{\pi 2} \parallel \left[r_{e1} + \frac{R_{sig}}{\beta_1 + 1} \right]$$

$$= 2500 \parallel \left[25 + \frac{4000}{101} \right] = 63\ \Omega$$

Capacitance $C_{\mu 2}$ sees a resistance $R_{\mu 2}$. To determine $R_{\mu 2}$ we refer to the analysis of the frequency response of the CE amplifier in Section 10.3.4 to obtain

$$R_{\mu 2} = (1 + g_{m2} R_L)(R_{in2} \parallel R_{out1}) + R_L$$

$$= (1 + 40 \times 4)\left[2500 \parallel \left(25 + \frac{4000}{101} \right) \right] + 4000$$

$$= 14{,}143\ \Omega\ = 14.1\ \text{k}\Omega$$

We now can determine τ_H from

$$\tau_H = C_{\mu 1} R_{\mu 1} + C_{\pi 1} R_{\pi 1} + C_{\mu 2} R_{\mu 2} + C_{\pi 2} R_{\pi 2}$$

$$= 2 \times 3.94 + 13.9 \times 0.0634 + 2 \times 14.1 + 13.9 \times 0.063$$

$$= 7.88 + 0.88 + 28.2 + 0.88 = 37.8\ \text{ns}$$

We observe that $C_{\pi 1}$ and $C_{\pi 2}$ play minor roles in determining the high-frequency response. As expected, $C_{\mu 2}$ through the Miller effect plays the most significant role. Capacitor $C_{\mu 1}$, which interacts directly with $(R_{sig} \parallel R_{in})$, also plays an important role. The 3-dB frequency f_H can be found as follows:

$$f_H = \frac{1}{2\pi \tau_H} = \frac{1}{2\pi \times 37.8 \times 10^{-9}} = 4.2\ \text{MHz}$$

Example 10.10 *continued*

For comparison, we evaluate A_M and f_H of a CE amplifier operating under the same conditions. Refer to Fig. 10.33(d). The voltage gain A_M is given by

$$A_M = \frac{R_{\text{in}}}{R_{\text{in}} + R_{\text{sig}}}(-g_m R_L) = \frac{r_\pi}{r_\pi + R_{\text{sig}}}(-g_m R_L)$$

$$= \frac{2.5}{2.5+4}(-40 \times 4) = -61.5 \text{ V/V}$$

$$R_\pi = r_\pi \| R_{\text{sig}} = 2.5 \| 4 = 1.54 \text{ k}\Omega$$

$$R_\mu = (1 + g_m R_L)(R_{\text{sig}} \| r_\pi) + R_L$$

$$= (1 + 40 \times 4)(4 \| 2.5) + 4$$

$$= 251.7 \text{ k}\Omega$$

Thus,

$$\tau_H = C_\pi R_\pi + C_\mu R_\mu$$

$$= 13.9 \times 1.54 + 2 \times 251.7$$

$$= 21.4 + 503.4 = 524.8 \text{ ns}$$

Observe the dominant role played by C_μ. The 3-dB frequency f_H is

$$f_H = \frac{1}{2\pi \tau_H} = \frac{1}{2\pi \times 524.8 \times 10^{-9}} = 303 \text{ kHz}$$

Thus, including the buffering transistor Q_1 increases the gain, $|A_M|$, from 61.5 V/V to 155 V/V—a factor of 2.5—and increases the bandwidth from 303 kHz to 4.2 MHz—a factor of 13.9! The gain–bandwidth product is increased from 18.63 MHz to 651 MHz—a factor of 35!

10.7.3 Increasing f_H by Eliminating the Miller Effect Using a CG or a CB Configuration with an Input Buffer

Our final approach to increasing f_H eliminates the Miller effect altogether by using a common-gate or a common-base amplifier. The drawback of the low input resistance of such amplifiers can be overcome by preceding them with a buffer (a source follower or an emitter follower). Three possible such configurations are depicted in Fig. 10.34. Since in each of these circuits, neither of the two transistors suffers from the Miller effect, the resulting

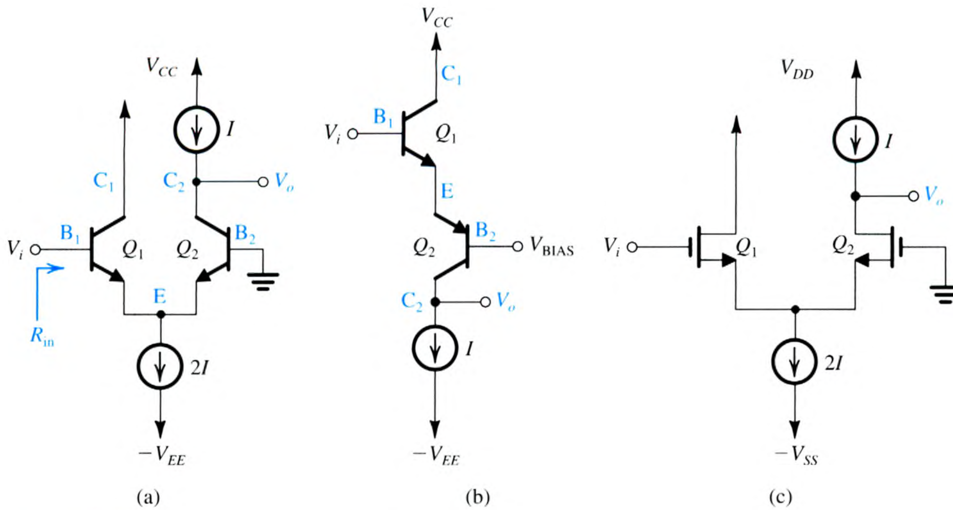


Figure 10.34 **(a)** A CC–CB amplifier. **(b)** Another version of the CC–CB circuit with Q_2 implemented using a *pnp* transistor. **(c)** The MOSFET version of the circuit in **(a)**. Notice that the circuits in **(a)** and **(c)** can be considered as differential amplifiers with the load resistance of Q_1 eliminated.

amplifiers have even wider bandwidths than those achieved in the circuits of the previous section. To illustrate, consider the circuit in Fig. 10.34(a).[7] The low-frequency analysis of this circuit gives for the input resistance,

$$R_{in} = (\beta_1 + 1)(r_{e1} + r_{e2}) \tag{10.144}$$

which for $r_{e1} = r_{e2} = r_e$ and $\beta_1 = \beta_2 = \beta$ becomes

$$R_{in} = 2r_\pi \tag{10.145}$$

If a load resistance R_L is connected at the output, the voltage gain V_o/V_i will be

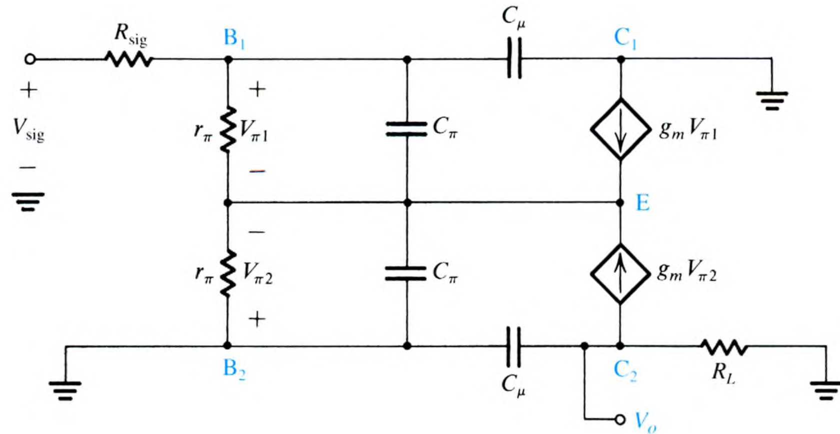
$$\frac{V_o}{V_i} = \frac{\alpha_2 R_L}{r_{e1} + r_{e2}} = \frac{1}{2} g_m R_L \tag{10.146}$$

where we have neglected r_o. Now, if the amplifier is fed with a voltage signal V_{sig} from a source with a resistance R_{sig}, the overall voltage gain will be

$$\frac{V_o}{V_{sig}} = \frac{1}{2}\left(\frac{R_{in}}{R_{in} + R_{sig}}\right)(g_m R_L) \tag{10.147}$$

The high-frequency analysis is illustrated in Fig. 10.35(a). Here we have drawn the hybrid-π equivalent circuit for each of Q_1 and Q_2. Recalling that the two transistors are operating at equal bias currents, their corresponding model components will be equal (i.e., $r_{\pi 1} = r_{\pi 2}$, $C_{\pi 1} = C_{\pi 2}$, etc.). With this in mind, you should be able to see (by writing a node equation at E) that $V_{\pi 1} = -V_{\pi 2}$ and the horizontal line through the node labeled E in Fig. 10.35(a) can be deleted. Thus the circuit reduces to that in Fig. 10.35(b). This is a

[7]The results derived for the circuit in Fig. 10.34(a) apply directly to the circuit of Fig. 10.34(b) and with appropriate change of variables to the MOS circuit of Fig. 10.34(c).



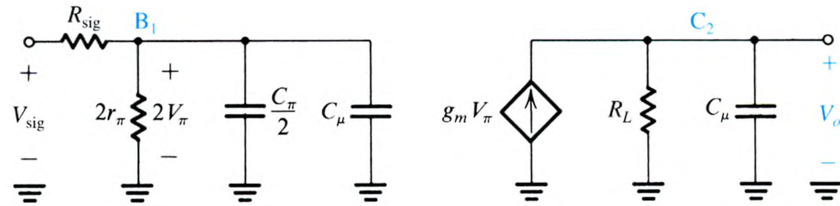


Figure 10.35 **(a)** Equivalent circuit for the amplifier in Fig. 10.34(a). **(b)** Simplified equivalent circuit. Note that the equivalent circuits in **(a)** and **(b)** also apply to the circuit shown in Fig. 10.34(b). In addition, they can be easily adapted for the MOSFET circuit in Fig. 10.34(c), with $2r_\pi$ eliminated, C_π replaced with C_{gs}, C_μ replaced with C_{gd}, and V_π replaced with V_{gs}.

very attractive outcome because the circuit shows clearly the two poles that determine the high-frequency response: The pole at the input, with a frequency f_{P1}, is

$$f_{P1} = \frac{1}{2\pi\left(\frac{C_\pi}{2} + C_\mu\right)(R_{sig} \| 2r_\pi)} \tag{10.148}$$

and the pole at the output, with a frequency f_{P2}, is

$$f_{P2} = \frac{1}{2\pi C_\mu R_L} \tag{10.149}$$

This result is also intuitively obvious: The input impedance at B_1 of the circuit in Fig. 10.35(a) consists of the series connection of $r_{\pi 1}$ and $r_{\pi 2}$ in parallel with the series connection of $C_{\pi 1}$ and $C_{\pi 2}$. Then there is $C_{\mu 1}$ in parallel. At the output, we simply have R_L in parallel with C_μ.

Whether one of the two poles is dominant will depend on the relative values of R_{sig} and R_L. If the two poles are close to each other, then the 3-dB frequency f_H can be determined either by exact analysis—that is, finding the frequency at which the gain is down by 3 dB—or

by using the approximate formula,

$$f_H \simeq 1 \Big/ \sqrt{\frac{1}{f_{P1}^2} + \frac{1}{f_{P2}^2}} \tag{10.150}$$

EXERCISE

10.25 For the CC–CB amplifier of Fig. 10.34(a), let $I = 0.5$ mA, $\beta = 100$, $C_\pi = 6$ pF, $C_\mu = 2$ pF, $R_{sig} = 10$ kΩ, and $R_L = 10$ kΩ. Find the low-frequency overall voltage gain A_M, the frequencies of the poles, and the 3-dB frequency f_H. Find f_H both exactly and using the approximate formula in Eq. (10.150).
Ans. 50 V/V; 6.4 MHz and 8 MHz; f_H by exact evaluation = 4.6 MHz; f_H using Eq. (10.150) = 5 MHz.

10.8 Low-Frequency Response of Discrete-Circuit CS and CE Amplifiers

So far in this chapter we have been concerned with the high-frequency response of amplifiers. The material presented applies equally well to both integrated-circuit and discrete-circuit amplifiers as both amplifier types behave similarly at high frequencies. This is not the case, however, at low frequencies (i.e., at frequencies lower than midband frequencies): While IC amplifiers exhibit a constant low-frequency gain equal to that at midband frequencies (Fig. 10.2), the gain of discrete-circuit amplifiers decreases at low frequencies and becomes zero at dc (Fig. 10.1). This is because discrete-circuit amplifiers use coupling and bypass capacitors. In this section we quantify the effect of such capacitors on the gain of discrete-circuit common-source and common-emitter amplifiers. Our goal is to be able to estimate the frequency f_L at which the gain drops by 3 dB below its value at midband and, of course, to design the amplifier so that it meets a specified value of f_L.

10.8.1 Frequency Response of the High-Pass Single-Time-Constant Circuit

Analyzing the low-frequency response of discrete-circuit amplifiers is much easier if we know the frequency response of the high-pass single-time-constant circuit shown in Fig. 10.36(a). The circuit has a time constant

$$\tau = CR \tag{10.151}$$

and a transfer function

$$\frac{V_o}{V_i} = \frac{s}{s + \omega_P} \tag{10.152}$$

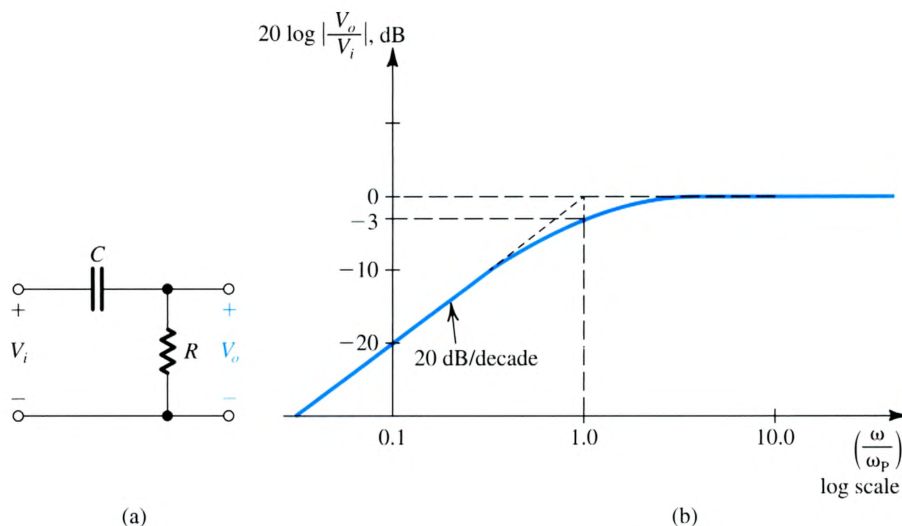


Figure 10.36 **(a)** High-pass STC circuit. **(b)** Bode plot for $|V_o/V_i|$.

where the pole-frequency ω_P is given by

$$\omega_P = \frac{1}{\tau} \tag{10.153}$$

In addition to the pole at $s = -\omega_P$, the transfer function in Eq. (10.152) has a zero at $s = 0$. That is, V_o/V_i becomes zero at $s = 0$. Physically, this is due to the capacitor C behaving as an open-circuit at $\omega = 0$ ($1/\omega C = \infty$ at $\omega = 0$) and thus causing $V_o = 0$.

For physical frequencies, $s = j\omega$, and the magnitude of the transfer function is

$$\left|\frac{V_o}{V_i}\right| = \frac{1}{\sqrt{1 + \left(\frac{\omega_P}{\omega}\right)^2}} \tag{10.154}$$

Thus, at $\omega \gg \omega_P$, $|V_o/V_i|$ approaches unity. At $\omega = \omega_P$, $|\frac{V_o}{V_i}| = 1/\sqrt{2}$; that is, the gain drops by 3 dB below its value at midband. Thus, the 3-dB frequency of the high-pass circuit in Fig. 10.36(a) is equal to the pole frequency ω_P.

For $\omega \ll \omega_P$,

$$\left|\frac{V_o}{V_i}\right| \simeq \frac{\omega}{\omega_P} \tag{10.155}$$

Thus, whenever ω decreases by a factor of 10 (a decade decrease), the gain decreases by a factor of 10 (i.e., by 20 dB). This is confirmed by the Bode plot for $|V_o/V_i|$ shown in Fig. 10.36(b). Finally, note that as expected, at $\omega = 0$ (dc), $|V_o/V_i| = 0$.

10.8.2 The CS Amplifier

Figure 10.37(a) shows a discrete-circuit common-source amplifier utilizing the classical biasing arrangement (Section 7.5.1). Two coupling capacitors, C_{C1} and C_{C2}, and a bypass capacitor C_S are used. At midband frequencies, these large capacitances have negligibly small

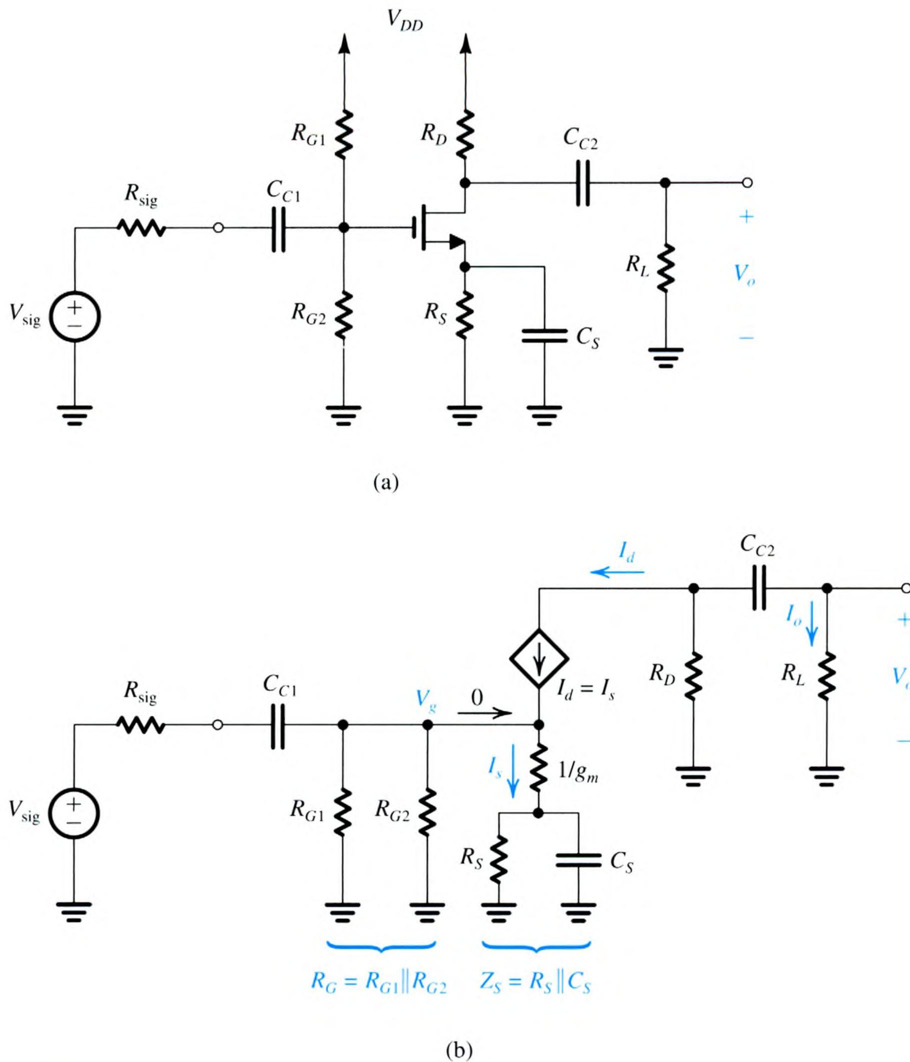


Figure 10.37 (**a**) Capacitatively coupled common-source amplifier. (**b**) The amplifier equivalent circuit at low frequencies. Note that the T model is used for the MOSFET and r_o is neglected.

impedances and can be assumed to be perfect short circuits for the purpose of calculating the midband gain, as was done in Section 7.5.1. However, at low frequencies, the reactance $1/j\omega C$ of each of the three capacitances increases and the amplifier gain decreases, as we will now show.

To determine the amplifier gain V_o/V_{sig} at low frequencies, we utilize the amplifier equivalent circuit shown in Fig. 10.37(b). This circuit is obtained by short-circuiting V_{DD} and replacing the MOSFET with its T model, which is the most convenient model to use when an impedance, such as Z_S, is present in the source lead. The transistor r_o has been omitted because, as mentioned in Chapter 7, r_o has a negligible effect on the performance of discrete-circuit amplifiers. Also, including r_o would complicate the analysis significantly.

The gain V_o/V_{sig} of the amplifier can be obtained from the equivalent circuit in Fig. 10.37(b) by starting at the source and working our way to the load, as

$$\frac{V_o}{V_{sig}} = \frac{V_g}{V_{sig}} \times \frac{I_d}{V_g} \times \frac{V_o}{I_d}$$

where V_g is the voltage between gate and ground, and I_d is the drain current. To find the fraction of V_{sig} that appears at the transistor gate, V_g, we note that the input resistance at the gate is infinite and thus the amplifier input resistance is $R_G = R_{G1} \| R_{G2}$. Using the voltage-divider rule gives

$$V_g = V_{sig} \frac{R_G}{R_G + \dfrac{1}{sC_{C1}} + R_{sig}}$$

which can be rearranged in the form

$$\frac{V_g}{V_{sig}} = \frac{R_G}{R_G + R_{sig}} \frac{s}{s + \dfrac{1}{C_{C1}(R_G + R_{sig})}} \tag{10.156}$$

Thus, we see that the effect of C_{C1} is to cause the expression for the signal transmission from the signal source to the amplifier input to acquire a frequency-dependent factor. We recognize this factor as the transfer function of a single-time-constant circuit of the high-pass type (Eq. 10.152), with a pole frequency ω_{P1},

$$\omega_{P1} = 1/C_{C1}(R_{sig} + R_G) \tag{10.157}$$

In addition to the pole, C_{C1} introduces a zero at $s = 0$ (dc). This is hardly surprising, since C_{C1} is included in the amplifier circuit to block dc. The Bode plot for the frequency-dependent factor in the transfer function of Eq. (10.156) will be similar to that in Fig. 10.36(b).

Continuing with the analysis, we next determine the drain current I_d, which is equal to the source current I_s. The latter can be found by dividing V_g by the total impedance in the source lead, $(1/g_m + Z_S)$,

$$I_d = I_s = \frac{V_g}{\dfrac{1}{g_m} + Z_S} = g_m V_g \frac{Y_S}{g_m + Y_S}$$

where

$$Y_S = \frac{1}{Z_S} = \frac{1}{R_S} + sC_S$$

Thus,

$$\frac{I_d}{V_g} = g_m \frac{s + \dfrac{1}{C_S R_S}}{s + \dfrac{g_m + 1/R_S}{C_S}} \tag{10.158}$$

Thus, the bypass capacitor introduces a pole with frequency ω_{P2},

$$\omega_{P2} = \frac{g_m + 1/R_S}{C_S} \tag{10.159}$$

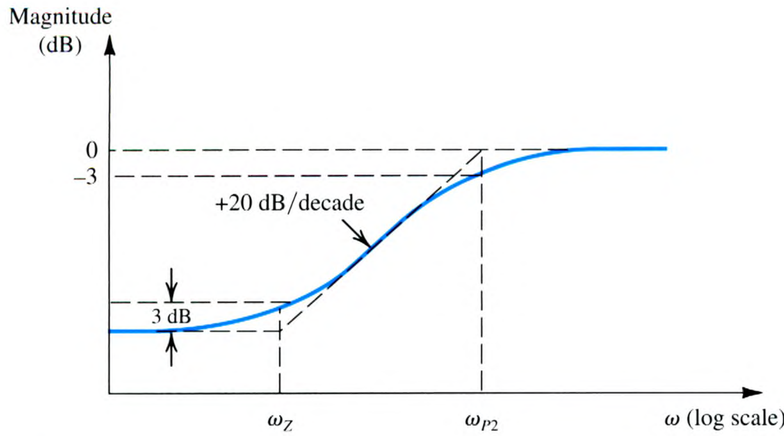


Figure 10.38 Sketch of the magnitude of the high-pass function $\dfrac{s+\omega_Z}{s+\omega_{P2}}$ versus frequency ω.

and a transmission zero on the negative real axis of the s plane at

$$s_Z = -\frac{1}{C_S R_S} \tag{10.160}$$

and thus has a frequency

$$\omega_Z = \frac{1}{C_S R_S} \tag{10.161}$$

Observe that since g_m is usually large, $\omega_{P2} \gg \omega_Z$. That is, ω_{P2} will be closer to the midband, and thus it plays a more significant role in determining ω_L than does ω_Z. Figure 10.38 shows a sketch of the magnitude of the frequency-dependent factor of the transfer function in Eq. (10.158).

To complete the analysis, we find V_o by first using the current-divider rule to determine the fraction of I_d that flows through R_L,

$$I_o = -I_d \frac{R_D}{R_D + \dfrac{1}{sC_{C2}} + R_L}$$

and then multiplying I_o by R_L. The result is

$$\frac{V_o}{I_d} = -\frac{R_D R_L}{R_D + R_L} \frac{s}{s + \dfrac{1}{C_{C2}(R_D + R_L)}} \tag{10.162}$$

from which we see that C_{C2} introduces a pole with frequency,

$$\omega_{P3} = \frac{1}{C_{C2}(R_D + R_L)} \tag{10.163}$$

and a zero at $s = 0$ (dc). A sketch of the magnitude of the frequency-dependent factor of the transfer function in Eq. (10.162) will be similar to that in Fig. 10.38.

The overall low-frequency gain function of the amplifier can be found by combining Eqs. (10.156), (10.158), and (10.162),

$$\frac{V_o}{V_{sig}} = -\frac{R_G}{R_G + R_{sig}} g_m (R_D \| R_L) \left(\frac{s}{s+\omega_{P1}}\right)\left(\frac{s+\omega_Z}{s+\omega_{P2}}\right)\left(\frac{s}{s+\omega_{P3}}\right)$$

$$\frac{V_o}{V_{sig}} = A_M \left(\frac{s}{s+\omega_{P1}}\right)\left(\frac{s+\omega_Z}{s+\omega_{P2}}\right)\left(\frac{s}{s+\omega_{P3}}\right) \tag{10.164}$$

where A_M, the midband gain, is given by

$$A_M = -\frac{R_G}{R_G + R_{sig}} g_m (R_D \| R_L) \tag{10.165}$$

which is the value we would have obtained had we assumed that C_{C1}, C_S, and C_{C2} were acting as perfect short circuits. In this regard, note from Eq. (10.164) that at midband frequencies—that is, at frequencies $s = j\omega$ with ω much higher than ω_{P1}, ω_{P2}, ω_Z, and ω_{P3}—V_o/V_{sig} approaches A_M, as should be expected.

Determining the 3-dB Frequency f_L The magnitude of the amplifier gain at a frequency ω can be obtained by substituting $s = j\omega$ in Eq. (10.164) and evaluating the magnitude of the resulting complex function. In this way, the low-frequency response of the amplifier can be plotted versus frequency, and the lower 3-dB frequency f_L can be determined as the frequency at which $|V_o/V_{sig}|$ drops to $|A_M|/\sqrt{2}$.

A simpler approach for determining f_L is possible if the poles and zeros are sufficiently separated from one another. In this case, we can employ the Bode plot rules (see Appendix F) to obtain a Bode plot for the gain magnitude. Such a plot is shown in Fig. 10.39. This graph is simply a combination of the Bode plots of the transfer functions in Eqs. (10.156), (10.158), and (10.162). Observe that since the poles and zeros are sufficiently separated, their effects appear distinct. As we move downward in frequency from the midband, we find that at each pole frequency, the slope of the asymptote to the gain function increases by 20 dB/decade, and at the zero frequency (f_Z) it decreases by 20 dB/decade. Note that for the purpose of this sketch, we assumed f_{P2} to be the highest of the three poles and zero frequencies, and that the zero has the lowest frequency.

A quick way for estimating the 3-dB frequency f_L is possible if the highest-frequency pole (here, assumed to be f_{P2}) is separated from the nearest pole or zero (here, f_{P1}) by at least a factor of 4 (two octaves). In such a case, f_L is approximately equal to the highest of the pole frequencies,

$$f_L \simeq f_{P2} \tag{10.166}$$

We refer to this situation as one in which a **dominant pole** exists, with the frequency of the dominant pole being f_{P2}. Of course, if a dominant pole exists, f_L can be estimated without the need for the Bode plot.

If a dominant pole does not exist, the following approximate expression for f_L can be used

$$f_L \simeq \sqrt{f_{P1}^2 + f_{P2}^2 + f_{P3}^2 - 2f_Z^2} \tag{10.167}$$

Determining the Pole and Zero Frequencies by Inspection From the derivation above of the transfer function V_o/V_{sig}, we note that each of the three capacitors determines the frequency of only one of the three poles and one of the three zeros. As a result of this isolation

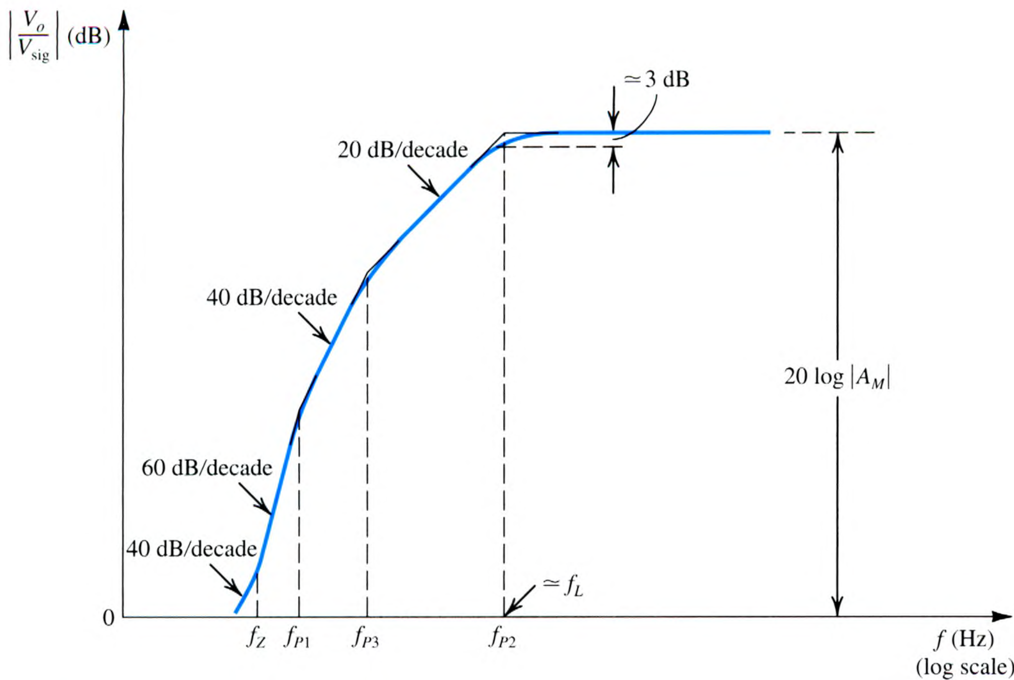


Figure 10.39 Sketch of the low-frequency magnitude response of a CS amplifier for which the three pole frequencies are sufficiently separated for their effects to appear distinct.

of the effects of the capacitors, a simple procedure exists for determining the frequencies of the pole and zero introduced by each capacitor.

Consider first the zeros. By its definition, a transmission zero is the value of s at which the input does not reach the output, resulting in $V_o = 0$. Examining the circuit in Fig. 10.37(b) shows us that C_{C1} becomes an infinite impedance at $s = 0$ and thus introduces a transmission zero at $s = 0$ (i.e., blocks dc).

An identical statement applies to C_{C2}. However, the bypass capacitor C_S has a different effect: Its transmission zero is at the value of s that causes Z_S to become infinite, and hence I_s, I_d, and V_o become zero, which is s_Z given by Eq. (10.160).

To determine the poles, we set $V_{sig} = 0$.[8] This results in the three separate circuits shown in Fig. 10.40. Each of the three circuits can be used to determine the resistance "seen" by the particular capacitor,[9] and hence the time constant associated with this capacitor. The corresponding pole frequency ω_P is the inverse of the time constant.

Selecting Values for the Coupling and Bypass Capacitors We now address the design issue of selecting appropriate values for C_{C1}, C_S, and C_{C2}. The design objective is to place the lower 3-dB frequency f_L at a specified value while minimizing the capacitor values. Since the resistance seen by C_S, $\left(\frac{1}{g_m} \parallel R_S\right)$, is usually the smallest of the three resistances,

[8]The poles of a circuit are its natural modes, and thus are independent of the value of the input signal.

[9]The resistance seen by a capacitor can be found by replacing the capacitor by a voltage source V_x and finding the current I_x drawn from V_x; the resistance is then V_x/I_x.

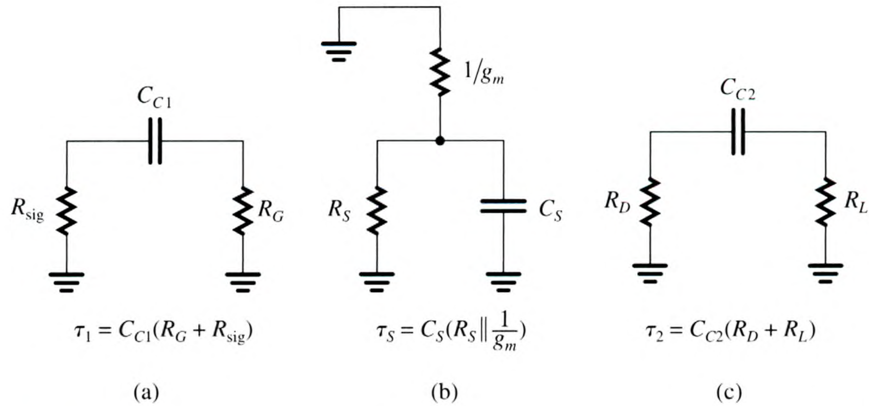


Figure 10.40 Circuits for determining the time constant of each of the three capacitors, and hence the pole associated with each one. Note that this determination is possible because in the circuit of Fig. 10.37, the capacitors do not interact; that is, each is responsible for only one of the three poles (and zeros).

the total capacitance is minimized by selecting C_S to provide the highest frequency pole; that is, making its pole frequency $f_{P2} = f_L$. We then decide on the location of the other two pole frequencies, say, 5 to 10 times lower than the frequency of the dominant pole, f_{P2}. However, the values selected for f_{P1} and f_{P3} should not be too low, for that would require larger values for C_{C1} and C_{C2} than may be necessary. The design procedure will be illustrated by an example.

Example 10.11

We wish to select appropriate values for the coupling capacitors C_{C1} and C_{C2} and the bypass capacitor C_S for a CS amplifier for which $R_G = 4.7\ \text{M}\Omega$, $R_D = R_L = 15\ \text{k}\Omega$, $R_{sig} = 100\ \text{k}\Omega$, $R_S = 10\ \text{k}\Omega$, and $g_m = 1\ \text{mA/V}$. We are required to have f_L at 100 Hz and the nearest break frequency at least a decade lower.

Solution

We select C_S so that

$$f_{P2} = \frac{g_m + \dfrac{1}{R_S}}{2\pi C_S} = f_L$$

Thus,

$$C_S = \frac{1.1 \times 10^{-3}}{2\pi \times 100} = 1.75\ \mu\text{F}$$

Then, selecting $f_{P1} = f_{P3} = 10$ Hz, we obtain

$$10 = \frac{1}{2\pi C_{C1}(0.1 + 4.7) \times 10^6}$$

which yields

$$C_{C1} = 3.3 \text{ nF}$$

and

$$10 = \frac{1}{2\pi C_{C2}(15+15)\times 10^3}$$

which results in

$$C_{C2} = 0.53 \ \mu\text{F}$$

Finally, we calculate the frequency of the zero f_Z as

$$f_Z = \frac{1}{2\pi C_S R_S} = \frac{1}{2\pi \times 1.75\times 10^{-6}\times 10\times 10^3}$$
$$= 9.1 \text{ Hz}$$

EXERCISE

10.26 A CS amplifier has $C_{C1} = C_S = C_{C2} = 1\ \mu\text{F}$, $R_G = 10\ \text{M}\Omega$, $R_{\text{sig}} = 100\,\text{k}\Omega$, $g_m = 2\ \text{mA/V}$, $R_D = R_L = R_S = 10\,\text{k}\Omega$. Find $A_M, f_{P1}, f_{P2}, f_{P3}, f_Z$, and f_L.
Ans. –9.9 V/V; 0.016 Hz; 334.2 Hz; 8 Hz; 15.91 Hz; 334.2 Hz

10.8.3 The Method of Short-Circuit Time Constants

In some circuits, such as that of the common-emitter amplifier discussed shortly, it is not possible to isolate the effects of the capacitors. Rather, the capacitors interact, making it difficult to determine the pole frequencies. Fortunately, there is a simple method for estimating f_L without having to determine the frequencies of the poles. Although the method is predicated on the assumption that one of the poles is dominant, the resulting estimate for f_L is usually very good even if this assumption is not strictly valid. The method is as follows:

1. Set the input signal $V_{\text{sig}} = 0$.
2. Consider the capacitors one at a time. That is, while considering capacitor C_i, set all the other capacitors to infinite values (i.e., replace them with short circuits—hence the name of the method).
3. For each capacitor C_i, find the total resistance R_i seen by C_i. This can be determined either by inspection or by replacing C_i with a voltage source V_x and finding the current I_x drawn from V_x; $R_i \equiv V_x/I_x$.

4. Calculate the 3-dB frequency f_L using

$$f_L \simeq \frac{1}{2\pi} \sum_{i=1}^{n} \frac{1}{C_i R_i} \tag{10.168}$$

where n is the total number of capacitors.

Besides its simplicity, this method has a very important side benefit: Equation (10.168) indicates the relative contribution of each capacitor to the value of f_L. Specifically, the lower the value of the time constant associated with a particular capacitor, the greater the contribution of this capacitor to f_L. As we will see shortly, this observation has important design implications. We will show how to apply the method of short-circuit time constants in the next section, where the method is used to determine f_L of the CE amplifier. Finally, note that this method is the dual of the method of open-circuit time constants studied in Section 10.3 for determining the upper 3-dB frequency f_H.

10.8.4 The CE Amplifier

Figure 10.41(a) shows a discrete-circuit common-emitter amplifier utilizing the classical biasing arrangement (Section 7.5.2), together with coupling capacitors C_{C1} and C_{C2}, and bypass capacitor C_E. We wish to obtain an estimate of the frequency f_L at which the gain of this amplifier drops by 3 dB below its value at midband. As well, we need to determine how to select appropriate values for C_{C1}, C_E, and C_{C2} to ensure that f_L is placed at a desired location while minimizing the total capacitance value required.

To analyze the low-frequency gain of the CE amplifier, we utilize the equivalent circuit shown in Fig. 10.41(b). This equivalent circuit is obtained by short-circuiting V_{CC} and replacing the BJT with its T model, while neglecting r_o. The decision to neglect r_o is based on the insignificant effect of the transistor's output resistance on the gain of discrete-circuit amplifiers, and the considerable complication its inclusion causes to the analysis. From the circuit in Fig. 10.41(b), we observe that the finite input current in the base of the BJT causes C_{C1} and C_{C2} to interact. That is, unlike the case of the CS amplifier, here each of the two poles caused by C_{C1} and C_{C2} will depend on both capacitor values in a complicated fashion that hinders design insight. Therefore, we will not attempt to determine the pole frequencies; instead, we will use the method of short-circuit time constants to estimate f_L directly.

Applying the Method of Short-Circuit Time Constants Setting $V_{sig} = 0$ in the circuit of Fig. 10.41(b) and considering each capacitor, one at a time, while short-circuiting the other two results in the three circuits shown in Fig. 10.42. These circuits can be used to determine the resistance seen by each capacitor and hence its effective time constant. For C_{C1} we use the circuit in Fig. 10.42(a) and note that r_π is the input resistance at the base when C_E is short-circuited. Capacitor C_{C1} sees a resistance R_{C1}, which can be found by inspection as

$$R_{C1} = (R_B \| r_\pi) + R_{sig} \tag{10.169}$$

and the time constant associated with C_{C1} becomes

$$\tau_{C1} = C_{C1} R_{C1} \tag{10.170}$$

For C_E, we use the circuit in Fig. 10.42(b). Here, we see that with C_{C1} shorted [refer to Fig. 10.41(b)], the resistance in the base becomes $(R_B \| R_{sig})$, which can be reflected to the

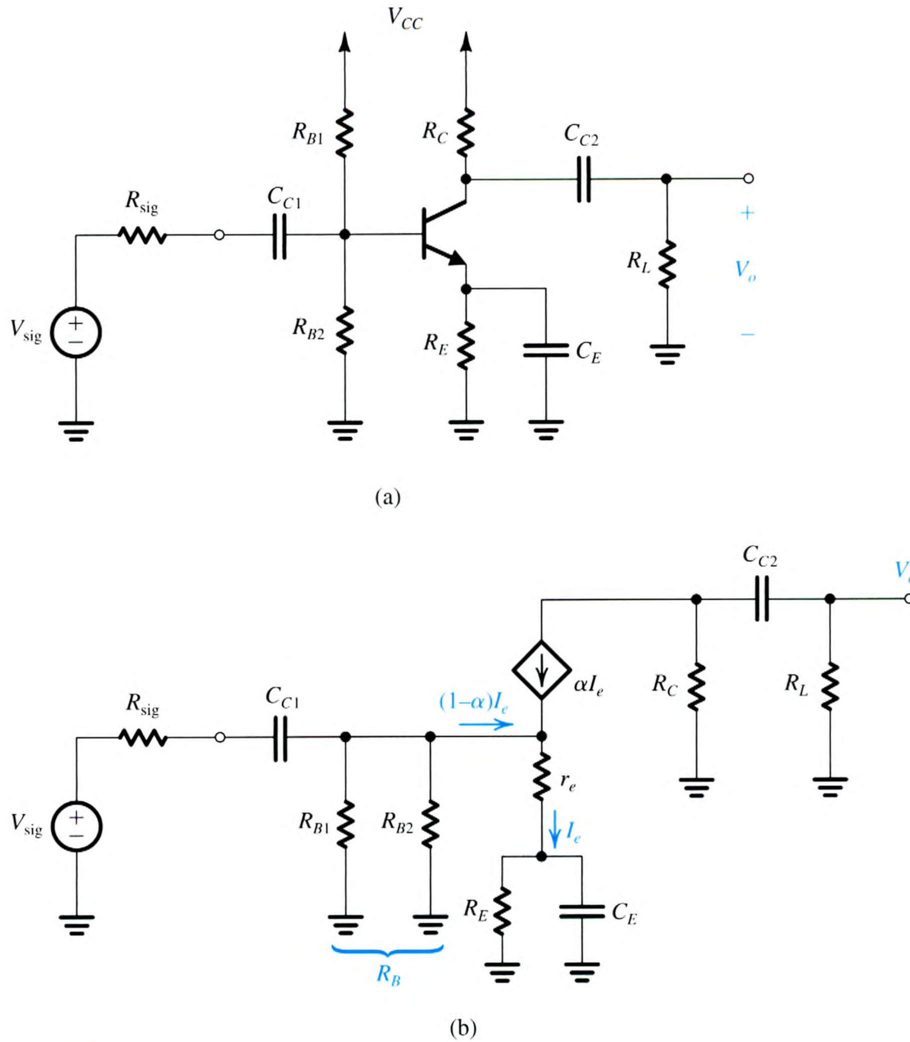


Figure 10.41 (**a**) A discrete-circuit common-emitter amplifier. (**b**) Equivalent circuit of the amplifier in (**a**).

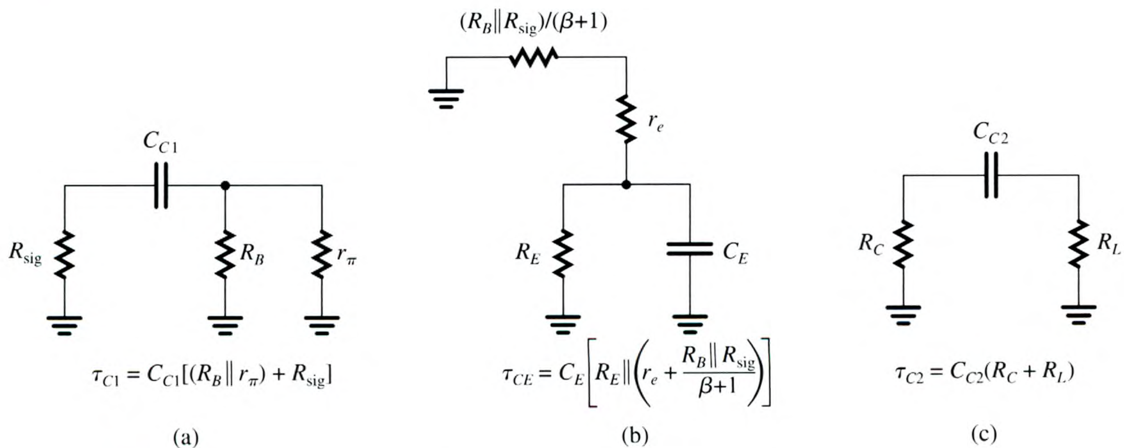


Figure 10.42 Circuits for determining the short-circuit time constants for the amplifier in Fig. 10.41.

emitter side as $(R_B \| R_{sig})/(\beta + 1)$. The total resistance R_{CE} seen by C_E can be found by inspection from the circuit in Fig. 10.42(b) as

$$R_{CE} = R_E \| \left[r_e + \frac{R_B \| R_{sig}}{\beta + 1} \right] \tag{10.171}$$

and the time constant becomes

$$\tau_{CE} = C_E R_{CE} \tag{10.172}$$

Finally, the resistance seen by C_{C2} can be determined by inspection of the circuit in Fig. 10.42(c) as

$$R_{C2} = R_C + R_L \tag{10.173}$$

and the corresponding time constant τ_{C2} as

$$\tau_{C2} = C_{C2} R_{C2} \tag{10.174}$$

With the three time constants in hand, the 3-dB frequency f_L can be found from

$$f_L = \frac{\omega_L}{2\pi} = \frac{1}{2\pi} \left[\frac{1}{C_{C1} R_{C1}} + \frac{1}{C_E R_{CE}} + \frac{1}{C_{C2} R_{C2}} \right] \tag{10.175}$$

When numerical values are substituted in this expression, it quickly becomes obvious which of the three capacitors is contributing the most to f_L. Obviously, it is the capacitor that has the smallest time constant. In the CE amplifier, this is usually C_E because the associated resistance R_{CE} is typically small. Knowing which of the capacitors has the potential of dominating the determination of f_L has significant design implications, as shown next.

Selecting Values for C_{C1}, C_E, and C_{C2} We now address the design issue of selecting appropriate values for C_{C1}, C_E, and C_{C2}. The design objective is to place the lower 3-dB frequency f_L at a specified location while minimizing the total capacitance. Since, as mentioned above, C_E usually sees the lowest of the three resistances, the total capacitance is minimized by selecting C_E so that its contribution to f_L is dominant. That is, by reference to Eq. (10.175) we may select C_E so that $1/(C_E R_{CE})$ is, say, 80% of $\omega_L = 2\pi f_L$, leaving each of the other capacitors to contribute 10% to the value of ω_L. Example 10.12 should help illustrate this process.

Example 10.12

We wish to select appropriate values for C_{C1}, C_{C2}, and C_E for the common-emitter amplifier, which has $R_B = 100$ kΩ, $R_C = 8$ kΩ, $R_L = 5$ kΩ, $R_{\text{sig}} = 5$ kΩ, $R_E = 5$ kΩ, $\beta = 100$, $g_m = 40$ mA/V, and $r_\pi = 2.5$ kΩ. It is required to have $f_L = 100$ Hz.

Solution

We first determine the resistances seen by the three capacitors C_{C1}, C_E, and C_{C2} as follows:

$$R_{C1} = (R_B \parallel r_\pi) + R_{\text{sig}}$$

$$= (100 \parallel 2.5) + 5 = 7.44 \text{ k}\Omega$$

$$R_{CE} = R_E \parallel \left[r_e + \frac{R_B \parallel R_{\text{sig}}}{\beta + 1} \right]$$

$$= 5 \parallel \left(0.025 + \frac{100 \parallel 5}{101} \right) = 0.071 \text{ k}\Omega$$

$$R_{C2} = R_C + R_L = 8 + 5 = 13 \text{ k}\Omega$$

Now, selecting C_E so that it contributes 80% of the value of ω_L gives

$$\frac{1}{C_E \times 71} = 0.8 \times 2\pi \times 100$$

$$C_E = 28 \ \mu\text{F}$$

Next, if C_{C1} is to contribute 10% of f_L,

$$\frac{1}{C_{C1} \times 7.44 \times 10^3} = 0.1 \times 2\pi \times 100$$

$$C_{C1} = 2.1 \ \mu\text{F}$$

Similarly, if C_{C2} is to contribute 10% of f_L, its value should be selected as follows:

$$\frac{1}{C_{C2} \times 13 \times 10^3} = 0.1 \times 2\pi \times 100$$

$$C_{C2} = 1.2 \ \mu\text{F}$$

In practice, we would select the nearest standard values for the three capacitors while ensuring that $f_L \leq$ 100 Hz. Finally, the frequency of the zero introduced by C_E can be found,

$$f_Z = \frac{1}{2\pi C_E R_E} = \frac{1}{2\pi \times 28 \times 10^{-6} \times 5 \times 10^3} = 1.1 \text{ Hz}$$

which is very far from f_L and thus has an insignificant effect.

EXERCISE

10.27 A common-emitter amplifier has $C_{C1} = C_E = C_{C2} = 1\,\mu\text{F}$, $R_B = 100\,\text{k}\Omega$, $R_{\text{sig}} = 5\,\text{k}\Omega$, $g_m = 40$ mA/V, $r_\pi = 2.5\,\text{k}\Omega$, $R_E = 5\,\text{k}\Omega$, $R_C = 8\,\text{k}\Omega$, and $R_L = 5\,\text{k}\Omega$. Find the value of the time constant associated with each capacitor, and hence estimate the value of f_L. Also compute the frequency of the transmission zero introduced by C_E and comment on its effect on f_L.
Ans. $\tau_{C1} = 7.44$ ms; $\tau_{CE} = 0.071$ ms; $\tau_{C2} = 13$ ms; $f_L = 2.28$ kHz; $f_Z = 31.8$ Hz, which is much smaller than f_L and thus has a negligible effect on f_L

Summary

- Both the MOSFET and the BJT have internal capacitive effects that can be modeled by augmenting the device hybrid-π model with capacitances. Usually at least two capacitances are needed: C_{gs} and C_{gd} (C_π and C_μ for the BJT). A figure of merit for the high-frequency operation of the transistor is the frequency f_T at which the short-circuit current gain of the CS (CE) transistor reduces to unity. For the MOSFET, $f_T = g_m/2\pi\left(C_{gs} + C_{gd}\right)$, and for the BJT, $f_T = g_m/2\pi\left(C_\pi + C_\mu\right)$. A summary of the high-frequency model parameters for the MOSFET is presented in Table 10.1 and the corresponding information for the BJT in Table 10.2; both may be found in the Summary Tables supplement found online at www.oup.com/he/sedra-smith8e.

- The internal capacitances of the MOSFET and the BJT cause the amplifier gain to fall off at high frequencies. An estimate of the amplifier bandwidth is provided by the frequency f_H at which the gain drops 3 dB below its value at midband, A_M. A figure of merit for the amplifier is the gain–bandwidth product $GB = A_M f_H$. Usually, it is possible to trade off gain for increased bandwidth, with GB remaining nearly constant. For amplifiers with a dominant pole with frequency f_H, the gain falls off at a uniform 20-dB/decade rate, reaching 0 dB at $f_t = GB$.

- The high-frequency response of the CS and CE amplifiers is severely limited by the Miller effect: The small capacitance C_{gd} $\left(C_\mu\right)$ is multiplied by a factor approximately equal to the gain from gate to drain (base to collector) $g_m R_L'$ and thus gives rise to a large capacitance at the amplifier input. The increased C_{in} interacts with the signal-source resistance R_{sig} and causes the amplifier gain to have a 3-dB frequency $f_H = 1/(2\pi R_{\text{sig}} C_{\text{in}})$.

- The method of open-circuit time constants provides a simple and powerful way to obtain a reasonably good estimate of the upper 3-dB frequency f_H. The capacitors that limit the high-frequency response are considered one at time with $V_{\text{sig}} = 0$ and all the other capacitances set to zero (open circuited). The resistance seen by each capacitance is determined, and the overall time constant τ_H is obtained by summing the individual time constants. Then f_H is found as $1/2\pi\tau_H$.

- The CG and CB amplifiers do *not* suffer from the Miller effect. Thus the cascode amplifier, which consists of a cascade of CS and CG stages (CE and CB stages), can be designed to obtain wider bandwidth than that achieved in the CS (CE) amplifier alone. The key, however, is to design the cascode so that the gain obtained in the CS (CE) stage is low.

- The source and emitter followers can have complex-conjugate poles. Thus, their frequency response is evaluated using the complete transfer function. Followers of both types exhibit wide bandwidths.

- The high-frequency response of the differential amplifier can be obtained by considering the differential and common-mode half-circuits. The CMRR falls off at a relatively low frequency determined by the output impedance of the bias current source.

- The high-frequency response of the current-mirror-loaded differential amplifier is complicated by the fact that there are two signal paths between input and output: a direct path and one through the current mirror. Nevertheless, the dominant pole is determined by the total capacitance at the output node, C_L.

- Combining two transistors in a way that eliminates or minimizes the Miller effect can result in a much wider bandwidth. Some such configurations are presented in Section 10.7.

- The coupling and bypass capacitors utilized in discrete-circuit amplifiers cause the amplifier gain to fall off at low frequencies. In the CS amplifier, the capacitors do not interact, and the frequencies of the

low-frequency poles can be estimated by considering each of these capacitors separately and determining the resistance seen by the capacitor. The highest-frequency pole is the one that determines the lower 3-dB frequency f_L. In the CE amplifier, the capacitors interact, and thus the poles cannot be easily determined. Rather, the method of short-circuit time constants can be used to obtain an estimate of the 3-dB frequency, f_L.

PROBLEMS

Computer Simulation Problems

SIM Problems identified by this icon are intended to demonstrate the value of using SPICE simulation to verify hand analysis and design, and to investigate important issues such as gain–bandwidth trade-off. Instructions to assist in setting up simulations for all the indicated problems can be found in the corresponding files on the website. Note that if a particular parameter value is not specified in the problem statement, you are to make a reasonable assumption.

Section 10.1: High-Frequency Transistor Models

10.1 Refer to the MOSFET high-frequency model in Fig. 10.4(a). Evaluate the model parameters for an NMOS transistor operating at $I_D = 200\ \mu\text{A}$, $V_{SB} = 0.35$ V, and $V_{DS} = 0.7$ V. The MOSFET has $W = 12\ \mu\text{m}$, $L = 0.3\ \mu\text{m}$, $t_{ox} = 4$ nm, $\mu_n = 450\ \text{cm}^2/\text{V}\cdot\text{s}$, $\gamma = 0.5\ \text{V}^{1/2}$, $2\phi_f = 0.85$ V, $\lambda = 0.1\ \text{V}^{-1}$, $V_0 = 0.9$ V, $C_{sb0} = C_{db0} = 5$ fF, and $L_{ov} = 0.03\ \mu\text{m}$. [Recall that $g_{mb} = \chi g_m$, where $\chi = \gamma/(2\sqrt{2\phi_f + V_{SB}})$, and that $\epsilon_{ox} = 3.45 \times 10^{-11}$ F/m.]

10.2 Find f_T for a MOSFET operating at $I_D = 200\ \mu\text{A}$ and $V_{OV} = 0.2$ V. The MOSFET has $C_{gs} = 20$ fF and $C_{gd} = 5$ fF.

10.3 Starting from the expression of f_T for a MOSFET,

$$f_T = \frac{g_m}{2\pi(C_{gs} + C_{gd})}$$

and making the approximation that $C_{gs} \gg C_{gd}$ and that the overlap component of C_{gs} is negligibly small, show that

$$f_T \simeq \frac{1.5}{\pi L}\sqrt{\frac{\mu_n I_D}{2C_{ox}WL}}$$

Thus note that to obtain a high f_T from a given device, it must be operated at a high current. Also note that faster operation is obtained from smaller devices.

10.4 Starting from the expression for the MOSFET unity-gain frequency,

$$f_T = \frac{g_m}{2\pi(C_{gs} + C_{gd})}$$

and making the approximation that $C_{gs} \gg C_{gd}$ and that the overlap component of C_{gs} is negligibly small, show that for an n-channel device

$$f_T \simeq \frac{3\mu_n V_{OV}}{4\pi L^2}$$

Observe that for a given channel length, f_T can be increased by operating the MOSFET at a higher overdrive voltage. Evaluate f_T for devices with $L = 0.2\ \mu\text{m}$ operated at overdrive voltages of 0.1 V and 0.2 V. Use $\mu_n = 450\ \text{cm}^2/\text{V}\cdot\text{s}$.

10.5 Find the intrinsic gain A_0 and the unity-gain frequency f_T of an n-channel transistor fabricated in a 0.13-μm CMOS process for which $L_{ov} = 0.1\ L$, $\mu_n = 400\ \text{cm}^2/\text{V}\cdot\text{s}$, and $V_A' = 5$ V/μm. The device is operated at $V_{OV} = 0.2$ V. Find A_0 and f_T for devices with $L = L_{\min}$, $2L_{\min}$, $3L_{\min}$, $4L_{\min}$, and $5L_{\min}$. Present your results in a table. (*Hint:* For f_T, use the approximate expression $f_T \simeq \dfrac{3\mu_n V_{OV}}{4\pi L^2}$.)

10.6 A particular BJT operating at $I_C = 0.5$ mA has $C_\mu = 0.5$ pF, $C_\pi = 5$ pF, and $\beta_0 = 100$. What are f_T and f_β for this situation?

---

problems with blue numbers are considered essential; * = difficult problem; ** = more difficult; *** = very challenging

SIM = simulation; **D** = design problem; ● = see related video example

Transistor	I_E(mA)	r_e(Ω)	g_m(mA/V)	r_π (kΩ)	β_0	f_T(MHz)	C_μ(pF)	C_π(pF)	f_β(MHz)
(a)	2				100	500	2		
(b)		25					2	10.7	4
(c)				2.5	100	500		10.7	
(d)	10				100	500	2		
(e)	0.1				100	150	2		
(f)	1				10	500	2		
(g)						800	1	9	80

10.7 For the transistor described in Problem 10.6, C_π includes a relatively constant depletion-layer capacitance of 1 pF. If the device is operated at $I_C = 0.25$ mA, what does its f_T become?

10.8 An *npn* transistor is operated at $I_C = 1$ mA and $V_{CB} = 2$ V. It has $\beta_0 = 100$, $V_A = 50$ V, $\tau_F = 30$ ps, $C_{je0} = 20$ fF, $C_{\mu 0} = 30$ fF, $V_{0c} = 0.75$ V, and $m_{CBJ} = 0.5$. Sketch the complete hybrid-π model, and specify the values of all its components. Also, find f_T.

10.9 Measurement of β of an *npn* transistor at 50 MHz shows that $|\beta| = 10$ at $I_C = 0.2$ mA and 12 at $I_C = 1.0$ mA. Furthermore, C_μ was measured and found to be 0.1 pF. Find f_T at each of the two collector currents used. What must τ_F and C_{je} be?

10.10 A particular small-geometry BJT has f_T of 12 GHz and $C_\mu = 0.05$ pF when operated at $I_C = 0.5$ mA. What is C_π in this situation? Also, find g_m. For $\beta_0 = 120$, find r_π and f_β.

10.11 For a BJT whose unity-gain bandwidth is 5 GHz and $\beta_0 = 200$, at what frequency does the magnitude of β become 50? What is f_β?

***10.12** Complete the table entries at the top of this page for transistors (a) through (g), under the conditions indicated.

Section 10.2: High-Frequency Response of CS and CE Amplifiers

10.13 In a multistage amplifier, the output resistance of stage 1 is found to be 10 kΩ. The succeeding stage, stage 2, has an input impedance that is mostly capacitive with a value of 10 pF. Find the frequency of the pole that arises as a result of the interconnection between these two amplifier stages.

10.14 A signal source modeled by its Norton's equivalent circuit has $R_{sig} = 1$ MΩ. It feeds an amplifier whose input impedance consists of a 1-MΩ resistance in parallel with a 0.1-pF capacitance. Find the frequency of the pole that arises as a result of this connection.

VE 10.1 **10.15** A common-source amplifier is fed with a signal source having a resistance $R_{sig} = 200$ kΩ and feeds a load resistance $R_L = 20$ kΩ. The MOSFET is operating at $g_m = 4$ mA/V and has $r_o = 20$ kΩ. The device capacitances are $C_{gs} = 25$ fF and $C_{gd} = 5$ fF. Find the midband gain A_M, the 3-dB frequency f_H, and the frequency of the transmission zero.

10.16 In a particular common-source amplifier for which the midband voltage gain between gate and drain (i.e., $-g_m R'_L$) is -29 V/V, the NMOS transistor has $C_{gs} = 1.0$ pF and $C_{gd} = 0.1$ pF. What input capacitance would you expect? For what range of signal-source resistances can you expect the 3-dB frequency to exceed 2 MHz?

D 10.17 In the circuit of Fig. P10.17, the voltage amplifier is ideal (i.e., it has an infinite input resistance and a zero output resistance).

(a) Use the Miller approach to find an expression for the input capacitance C_{in} in terms of A and C.
(b) Use the expression for C_{in} to obtain the transfer function $V_o(s)/V_{sig}(s)$.

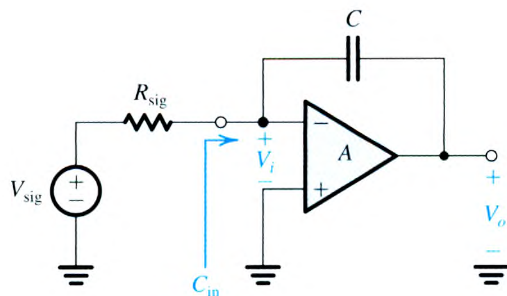


Figure P10.17

(c) If $R_{sig} = 1$ kΩ, and the gain V_o/V_{sig} is to have a dc value of 40 dB and a 3-dB frequency of 100 kHz, find the values required for A and C.

(d) Sketch a Bode plot for the gain and use it to determine the frequency at which its magnitude reduces to unity.

10.18 An ideal voltage amplifier having a voltage gain of −1000 V/V has a 0.1-pF capacitance connected between its output and input terminals. What is the input capacitance of the amplifier? If the amplifier is fed from a voltage source V_{sig} having a resistance $R_{sig} = 10$ kΩ, find the transfer function V_o/V_{sig} as a function of the complex-frequency variable s and hence the 3-dB frequency f_H and the unity-gain frequency f_t.

D 10.19 Design a CS amplifier for which the MOSFET is operated at $g_m = 5$ mA/V and has $C_{gs} = 5$ pF and $C_{gd} = 1$ pF. The amplifier is fed with a signal source having $R_{sig} = 1$ kΩ. What is the largest value of R_L' for which the upper 3-dB frequency is at least 6 MHz? What is the corresponding value of midband gain and gain–bandwidth product? If the specification on the upper 3-dB frequency can be relaxed by a factor of 3, that is, to 2 MHz, what can A_M and GB become?

10.20 Reconsider Example 10.1 for the situation in which the transistor is replaced by one whose width W is half that of the original transistor while the bias current remains unchanged. Find modified values for all the device parameters along with A_M, f_H, and the gain–bandwidth product, GB. Contrast this with the original design by calculating the ratios of new value to old for W, V_{OV}, g_m, C_{gs}, C_{gd}, C_{in}, A_M, f_H, and GB.

D *10.21 In a CS amplifier, the resistance of the signal source $R_{sig} = 100$ kΩ, amplifier input resistance (which is due to the biasing network) $R_{in} = 100$ kΩ, $C_{gs} = 1$ pF, $C_{gd} = 0.2$ pF, $g_m = 5$ mA/V, $r_o = 25$ kΩ, and $R_L = 20$ kΩ. Determine the expected 3-dB cutoff frequency f_H and the midband gain. In evaluating ways to double f_H, a designer considers the alternatives of changing either R_L or R_{in}. To raise f_H as described, what separate change in each would be required? What midband voltage gain results in each case?

SIM **10.22** Consider the integrated-circuit CS amplifier in Fig. P10.22 for the case $I_{BIAS} = 100$ μA, Q_2 and Q_3 are matched, and $R_{sig} = 20$ kΩ. For Q_1: $\mu_n C_{ox} = 400$ μA/V^2, $V_A = 5$ V, $W/L = 5$ μm/0.4 μm, $C_{gs} = 30$ fF, and $C_{gd} = 5$ fF. For Q_2: $|V_A| = 5$ V. Neglecting the effect of the capacitance inevitably present at the output node, find the low-frequency gain, the 3-dB frequency f_H, and the frequency of the zero f_Z.

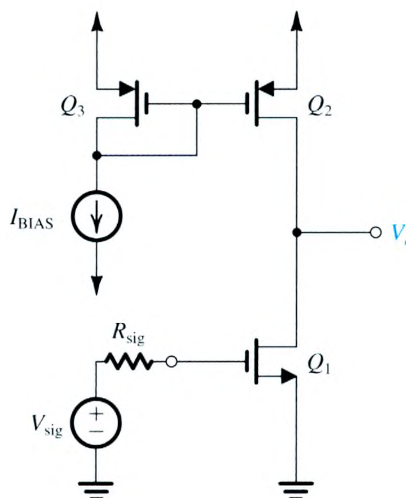


Figure P10.22

10.23 An IC CS amplifier is fed from a signal source with a negligibly small resistance and has a total effective load resistance $R_L' = 20$ kΩ. The MOSFET is operating at $g_m = 2$ mA/V and has a $C_{gd} = 10$ fF. The total capacitance C_L at the output node is 100 fF. Find the midband gain A_M, the 3-dB frequency f_H, the unity-gain frequency f_t, the frequency of the transmission zero f_z, and the gain at very-high frequencies. Sketch and clearly label the Bode plot for the gain magnitude.

10.24 An IC CS amplifier fed with a signal source of a negligibly small resistance R_{sig} is found to have a low-frequency gain of −50 V/V. The gain falls off with frequency at −20 dB/decade and reaches zero dB at 1 GHz. The gain levels off at higher frequencies at −20 dB. If the MOSFET is operating with $g_m = 2$ mA/V, estimate the values of C_{gd} and C_L. Also give the value of the 3-dB frequency f_H.

D 10.25 A common-source amplifier fed with a low-resistance signal source and operating with $g_m = 2$ mA/V

has a unity-gain frequency of 2 GHz. What additional capacitance must be connected to the drain node to reduce f_t to 1 GHz?

10.26 Analyze the high-frequency response of the CMOS amplifier shown in Fig. P10.22 for the case $R_{sig} = 0$. The dc bias current is 100 μA. For Q_1, $\mu_n C_{ox} = 400\ \mu A/V^2$, $V_A = 5$ V, $W/L = 5\ \mu m/0.4\ \mu m$, $C_{gs} = 30$ fF, $C_{gd} = 5$ fF, and $C_{db} = 5$ fF. For Q_2, $C_{gd} = 5$ fF, $C_{db} = 10$ fF, and $|V_A| = 5$ V. For simplicity, assume that the signal voltage at the gate of Q_2 is zero. Find the low-frequency gain, the frequency of the pole, and the frequency of the zero. (*Hint:* The total capacitance at the output mode $= C_{db1} + C_{db2} + C_{gd2}$.)

10.27 A common-emitter amplifier is measured at midband and found to have a gain of −50 V/V between base and collector. If $C_\pi = 10$ pF, $C_\mu = 1$ pF, and the effective source resistance $R'_{sig} = 5\ k\Omega$ [refer to Fig. 10.13(c)], find C_{in} and the 3-dB frequency f_H.

10.28 For a CE amplifier represented by the equivalent circuit in Fig. 10.13(b), let $R_{sig} = 10\ k\Omega$, $C_\pi = 10$ pF, $C_\mu = 0.5$ pF, $g_m = 40$ mA/V, $r_o = 100\ k\Omega$, $R_L = 10\ k\Omega$, and $\beta = 100$. Find the midband gain and the 3-dB frequency f_H.

***10.29** We want to investigate the high-frequency response of the CE amplifier when it is fed with a relatively large source resistance R_{sig}. Refer to the amplifier in Fig. 10.13(a) and to its high-frequency, equivalent-circuit model and the analysis shown. Let $R_{sig} \gg r_\pi$, $g_m R'_L \gg 1$, and $g_m R'_L C_\mu \gg C_\pi$. Under these conditions, show that:

(a) the midband gain $A_M \simeq -\beta R'_L / R_{sig}$
(b) the upper 3-dB frequency $f_H \simeq 1/2\pi C_\mu \beta R'_L$
(c) the gain–bandwidth product $|A_M| f_H \simeq 1/2\pi C_\mu R_{sig}$

Evaluate this approximate value of the gain–bandwidth product if $R_{sig} = 25\ k\Omega$ and $C_\mu = 1$ pF. Now, if the transistor has $\beta = 100$, find the midband gain and f_H for the two cases $R'_L = 25\ k\Omega$ and $R'_L = 2.5\ k\Omega$. On the same coordinates, sketch Bode plots for the gain magnitude versus frequency for the two cases. What f_H is obtained when the gain is unity? What value of R'_L corresponds?

10.30 The discrete-circuit common-emitter amplifier in Fig. P10.30 has $R_{sig} = 10\ k\Omega$, $R_{B1} = 68\ k\Omega$, $R_{B2} = 27\ k\Omega$, $R_E = 2.2\ k\Omega$, $R_C = 4.7\ k\Omega$, and $R_L = 10\ k\Omega$. The transistor is operating at a dc collector current of 0.8 mA and has $\beta = 160$, $f_T = 1$ GHz, and $C_\mu = 0.5$ pF. Give the high-frequency small-signal equivalent-circuit model of the amplifier assuming the coupling and bypass capacitors behave as short circuits, and neglecting r_o. Find A_M and f_H.

***10.31** Figure P10.31 shows a diode-connected transistor with the bias circuit omitted. Utilizing the BJT high-frequency, hybrid-π model with $r_o = \infty$, derive an

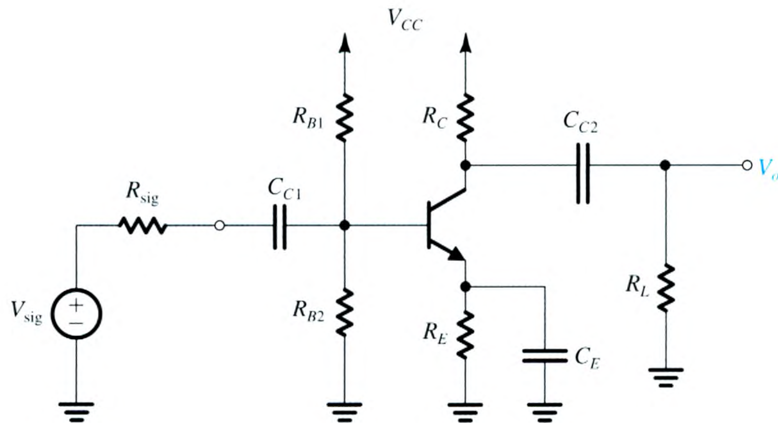


Figure P10.30

expression for $Z_i(s)$ as a function of r_e and C_π. Find the frequency at which the impedance has a phase angle of 45° for the case in which the BJT has $f_T = 400$ MHz and the bias current is relatively high, so that $C_\pi \gg C_\mu$. What is the frequency when the bias current is reduced so that $C_\pi \simeq C_\mu$? Assume $\alpha = 1$.

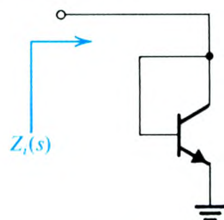


Figure P10.31

10.32 Consider a current-source-loaded common-emitter amplifier. Let the amplifier be fed with an ideal voltage source V_i. Assume that the load current source has a very high resistance and that there is a capacitance C_L present between the output node and ground. This capacitance represents the sum of the input capacitance of the subsequent stage and the inevitable parasitic capacitance between collector and ground. Show that the voltage gain is given by

$$\frac{V_o}{V_i} = -g_m r_o \frac{1 - s(C_\mu/g_m)}{1 + s(C_L + C_\mu)r_o}$$

If the transistor is biased at $I_C = 200\ \mu$A and $V_A = 50$ V, $C_\mu = 0.2$ pF, and $C_L = 1$ pF, find the dc gain, the 3-dB frequency, the frequency of the zero, and the frequency at which the gain reduces to unity. Sketch a Bode plot for the gain magnitude.

10.33 Consider an ideal voltage amplifier with a gain of 0.9 V/V, and a resistance R = 100 kΩ connected in the feedback path—that is, between the output and input terminals. Use Miller's theorem to find the input resistance of this circuit.

10.34 The amplifiers listed below are characterized by the descriptor (A, C), where A is the voltage gain from input to output and C is a capacitor connected between input and output. For each, find the equivalent capacitances at the input and at the output as provided by the use of Miller's theorem:

(a) –1000 V/V, 1 pF
(b) –10 V/V, 10 pF
(c) –1 V/V, 10 pF
(d) +1 V/V, 10 pF
(e) +10 V/V, 10 pF

Note that the input capacitance found in case (e) can be used to cancel the effect of other capacitance connected from input to ground. In (e), what capacitance can be canceled?

***10.35** Figure P10.35 shows an ideal voltage amplifier with a gain of +2 V/V (usually implemented with an op amp connected in the noninverting configuration) and a resistance R connected between output and input.

(a) Using Miller's theorem, show that the input resistance $R_{in} = -R$.
(b) Use Norton's theorem to replace V_{sig}, R_{sig}, and R_{in} with a signal current source and an equivalent parallel resistance. Show that by selecting $R_{sig} = R$, the equivalent parallel resistance becomes infinite and the current I_L into the load impedance Z_L becomes V_{sig}/R. The circuit then functions as an ideal voltage-controlled current source with an output current I_L.
(c) If Z_L is a capacitor C, find the transfer function V_o/V_{sig} and show it is that of an ideal noninverting integrator.

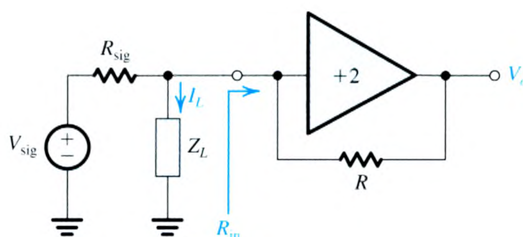


Figure P10.35

10.36 Use Miller's theorem to investigate the performance of the inverting op-amp circuit shown in Fig. P10.36. Assume the op amp to be ideal except for having a finite differential gain, A. Without using any knowledge of op-amp

A	R_{in}	V_i	V_o	V_o/V_{sig}
10 V/V				
100 V/V				
1000 V/V				
10,000 V/V				

circuit analysis, find R_{in}, V_i, V_o, and V_o/V_{sig}, for each of the following values of A: 10 V/V, 100 V/V, 1000 V/V, and 10,000 V/V. Assume V_{sig} = 1 V. Present your results in the table on page 769.

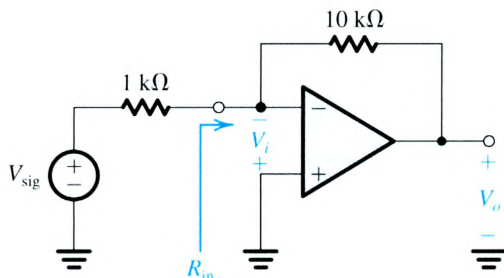


Figure P10.36

***10.37** The amplifier shown in Fig. P10.37 has $R_{sig} = R_L = 1\text{ k}\Omega$, $R_C = 1\text{ k}\Omega$, $R_B = 47\text{ k}\Omega$, $\beta = 100$, $C_\mu = 0.8$ pF, and $f_T = 600$ MHz. Assume the coupling capacitors to be very large.

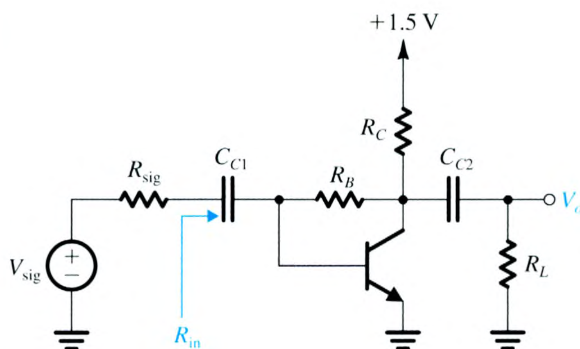


Figure P10.37

(a) Find the dc collector current of the transistor.
(b) Find g_m and r_π.
(c) Neglecting r_o, find the midband voltage gain from base to collector (neglect the effect of R_B).
(d) Use the gain obtained in (c) to find the component of R_{in} that arises as a result of R_B. Hence find R_{in}.
(e) Find the overall gain at midband.
(f) Find C_{in}.
(g) Find f_H.

Section 10.3: The Method of Open-Circuit Time Constants

10.38 A direct-coupled amplifier has a low-frequency gain of 40 dB, poles at 1 MHz and 10 MHz, a zero on the negative real axis at 100 MHz, and another zero at infinite frequency. Express the amplifier gain function in the form of Eqs. (10.56) and (10.57), and sketch a Bode plot for the gain magnitude. Does this amplifier have a dominant pole? If so, at what frequency? What do you estimate the 3-dB frequency f_H to be?

10.39 An amplifier with a dc gain of 60 dB has a single-pole, high-frequency response with a 3-dB frequency of 100 kHz.

(a) Give an expression for the gain function $A(s)$.
(b) Sketch a Bode diagram for the gain magnitude.
(c) What is the gain–bandwidth product?
(d) What is the unity-gain frequency?
(e) If a change in the amplifier circuit causes its transfer function to acquire another pole at 1 MHz, sketch the resulting gain magnitude and specify the unity-gain frequency. Note that this is an example of an amplifier with a unity-gain bandwidth that is different from its gain–bandwidth product.

10.40 An IC CS amplifier has $g_m = 3$ mA/V, $C_{gs} = 25$ fF, $C_{gd} = 5$ fF, $C_L = 30$ fF, $R_{sig} = 10\text{ k}\Omega$, and $R_L' = 20\text{ k}\Omega$. Use the method of open-circuit time constants to obtain an estimate for f_H. Also, find the frequency of the transmission zero, f_Z.

10.41 A discrete-circuit CS amplifier is modeled by the circuit of Fig. 10.16 augmented with a bias-resistance R_G between gate and ground. If $g_m = 5$ mA/V, $R_{sig} = 100\text{ k}\Omega$, $R_G = 500\text{ k}\Omega$, $R_L' = 10\text{ k}\Omega$, $C_{gs} = 1$ pF, $C_{gd} = 0.2$ pF, and $C_L = 20$ pF, find the midband gain and estimate f_H using the method of open-circuit time constants.

10.42 Consider the high-frequency response of an amplifier consisting of two identical transconductance amplifier stages in cascade, each with an input resistance of 10 kΩ and an output resistance of 2 kΩ. The two-stage amplifier is driven from a 10-kΩ source and drives a 1-kΩ load. Associated with

each stage is a parasitic input capacitance (to ground) of 10 pF and a parasitic output capacitance (to ground) of 2 pF. Parasitic capacitances of 10 pF and 7 pF also are associated with the signal-source and load connections, respectively. For this circuit, find the three poles and estimate the 3-dB frequency f_H.

10.43 A CS amplifier that can be represented by the equivalent circuit of Fig. 10.16 has $C_{gs} = 2$ pF, $C_{gd} = 0.1$ pF, $C_L = 2$ pF, $g_m = 4$ mA/V, and $R_{sig} = R_L' = 20$ kΩ. Find the midband gain A_M, the input capacitance C_{in} using the Miller effect, and hence an estimate of the 3-dB frequency f_H. Also, obtain another estimate of f_H using open-circuit time constants. Which of the two estimates is more appropriate and why?

D 10.44 For a CS amplifier with $g_m = 5$ mA/V, $C_{gs} = 5$ pF, $C_{gd} = 1$ pF, $C_L = 5$ pF, $R_{sig} = 10$ kΩ, and $R_L' = 10$ kΩ, find τ_H and f_H. What is the percentage of τ_H that is caused by the interaction of R_{sig} with the input capacitance? To what value must R_{sig} be lowered in order to double f_H?

D 10.45 For the CS amplifier in Example 10.5, find the value of the additional capacitance to be connected at the output node in order to lower f_H to 40 MHz.

10.46 Use the method of open-circuit time constants to find f_H for a CS amplifier for which $g_m = 1.5$ mA/V, $C_{gs} = C_{gd} = 0.2$ pF, $r_o = 20$ kΩ, $R_L = 12$ kΩ, and $R_{sig} = 100$ kΩ for the following cases: (a) $C_L = 0$, (b) $C_L = 10$ pF, and (c) $C_L = 50$ pF. Compare with the value of f_H obtained using the Miller effect.

10.47 Consider a CS amplifier loaded in a current source with an output resistance equal to r_o of the amplifying transistor. The amplifier is fed from a signal source with $R_{sig} = r_o/2$. The transistor is biased to operate at $g_m = 2$ mA/V and $r_o = 20$ kΩ; $C_{gs} = C_{gd} = 0.1$ pF. Use the Miller effect to determine an estimate of f_H. Repeat for the following two cases: (i) the bias current I in the entire system is reduced by a factor of 4, and (ii) the bias current I in the entire system is increased by a factor of 4. Remember that both R_{sig} and R_L will change as r_o changes.


VE 10.2

10.48 A common-emitter amplifier has $C_\pi = 10$ pF, $C_\mu = 0.3$ pF, $C_L = 3$ pF, $g_m = 40$ mA/V, $\beta = 100$, $R_L' = 5$ kΩ, and $R_{sig} = 1$ kΩ. Find the midband gain A_M and an estimate of the 3-dB frequency f_H using the Miller effect. Also, obtain another estimate of f_H using the method of open-circuit time constants. Which of the two estimates would you consider to be more realistic, and why?

10.49 Consider the CE amplifier whose equivalent circuit is shown in Fig. 10.13(b) but with a capacitance C_L connected across the output terminals. Let $R_{sig} = 5$ kΩ, $g_m = 20$ mA/V, $\beta = 100$, $C_\pi = 10$ pF, $C_\mu = 1$ pF, $R_L' = 5$ kΩ, and $C_L = 10$ pF. Find A_M and f_H.

Section 10.4: High-Frequency Response of Common-Gate and Cascode Amplifiers

10.50 A discrete-circuit CG amplifier is specified to have $C_{gs} = 5$ pF, $C_{gd} = 0.2$ pF, $C_L = 3$ pF, $g_m = 4$ mA/V, $R_{sig} = 1$ kΩ, and $R_L = 10$ kΩ. Neglecting the effects of r_o, find the low-frequency gain V_o/V_{sig}, the frequencies of the poles f_{P1} and f_{P2}, and hence an estimate of the 3-dB frequency f_H.

10.51 Sketch the high-frequency equivalent circuit of a CB amplifier fed from a signal generator characterized by V_{sig} and R_{sig} and feeding a load resistance R_L in parallel with a capacitance C_L. Neglect r_o.

(a) Show that the circuit can be separated into two parts: an input part that produces a pole at

$$f_{P1} = \frac{1}{2\pi C_\pi (R_{sig} \| r_e)}$$

and an output part that forms a pole at

$$f_{P2} = \frac{1}{2\pi (C_\mu + C_L) R_L}$$

Note that these are the bipolar counterparts of the MOS expressions in Eqs. (10.76) and (10.77).

(b) Evaluate f_{P1} and f_{P2} and hence obtain an estimate for f_H for the case $C_\pi = 10$ pF, $C_\mu = 1$ pF, $C_L = 1$ pF, $I_C = 1$ mA, $R_{sig} = 1$ kΩ, and $R_L = 10$ kΩ. Also, find f_T of the transistor.

10.52 Consider a CG amplifier loaded in a resistance $R_L = r_o$ and fed with a signal source having a resistance $R_{sig} = r_o/2$. Also let $C_L = C_{gs}$. Use the method of open-circuit time constants to show that for $g_m r_o \gg 1$, the upper 3-dB frequency is related to the MOSFET f_T by the approximate expression

$$f_H = f_T / (g_m r_o)$$

10.53 For the CG amplifier in Example 10.6, how much additional capacitance should be connected between the output node and ground to reduce f_H to 180 MHz?

10.54 An IC CG amplifier is fed from a signal source with $R_{sig} = r_o/2$, where r_o is the MOSFET output resistance. It has a current-source load with an output resistance equal

to r_o. The MOSFET is operated at $I_D = 100\ \mu A$ and has $g_m = 2$ mA/V, $V_A = 5$ V, $C_{gs} = 0.1$ pF, $C_{gd} = 10$ fF, and $C_{db} = 10$ fF. As well, the current-source load provides an additional 30 fF capacitance at the output node. Find f_H.

10.55 Consider a CG amplifier driven by a current-source signal I_{sig} having a predominantly capacitive source impedance; that is, the signal source can be represented by a current source I_{sig} in parallel with a capacitance C_{sig}. Neglecting r_o of the MOSFET, find an expression for the transfer function V_o/I_{sig} obtained when the load consists of a resistance R_L in parallel with a capacitance C_L. Hence find expressions for the frequencies of the two poles.



VE 10.3 **10.56** Find the dc gain and the 3-dB frequency of a MOS cascode amplifier operated at $g_m = 2$ mA/V and $r_o = 20$ kΩ. The MOSFETs have $C_{gs} = 20$ fF, $C_{gd} = 5$ fF, and $C_{db} = 5$ fF. The amplifier is fed from a signal source with $R_{sig} = 100$ kΩ and is connected to a load resistance of 1 MΩ. There is also a load capacitance C_L of 20 fF.

SIM **10.57** (a) Consider a CS amplifier having $C_{gd} = 0.3$ pF, $R_{sig} = R_L = 20$ kΩ, $g_m = 4$ mA/V, $C_{gs} = 2$ pF, C_L (including C_{db}) = 1 pF, $C_{db} = 0.2$ pF, and $r_o = 20$ kΩ. Find the low-frequency gain A_M, and estimate f_H using open-circuit time constants. Hence determine the gain–bandwidth product.

(b) If a CG stage utilizing an identical MOSFET is cascaded with the CS transistor in (a) to create a cascode amplifier, determine the new values of A_M, f_H, and gain–bandwidth product. Assume R_L remains unchanged.



VE 10.4 **D 10.58** Design a cascode amplifier to provide a dc gain of 74 dB when driven with a low-resistance generator and utilizing NMOS transistors for which $V_A = 10$ V, $\mu_n C_{ox} = 200\ \mu A/V^2$, $W/L = 50$, $C_{gd} = 0.1$ pF, and $C_L = 1$ pF. Assuming that $R_L = R_o$, determine the overdrive voltage and the drain current at which the MOSFETs should be operated. Find the unity-gain frequency and the 3-dB frequency. If the cascode transistor is removed and R_L remains unchanged, what will the dc gain, the 3-dB frequency, and the unity-gain frequency become?

10.59 (a) Show that introducing a cascode transistor to an IC CS amplifier whose bandwidth is limited by the interaction of R_{sig} and the input capacitance, and whose load resistance is equal to r_o, increases the dc gain by approximately a factor of 2 and f_H by the factor N,

$$N = \frac{C_{gs} + \frac{1}{2}(g_m r_o)C_{gd}}{C_{gs} + 3C_{gd}}$$

Assume that the bandwidth of the cascode amplifier is primarily determined by the input circuit.

(b) If $C_{gd} = 0.1\,C_{gs}$ and the dc gain of the CS amplifier is 50, what is the value of N?

(c) If $V_A = 10$ V, $\mu_n C_{ox} = 400\ \mu A/V^2$, and $W/L = 10$, find V_{OV} and I_D at which the transistors must be operating.

10.60 (a) For an integrated-circuit MOS cascode amplifier fed with a source having a very small resistance and loaded in a resistance equal to its R_o, use the expression for the unity-gain bandwidth in Fig. 10.21 to show that

$$f_t = \frac{\sqrt{2\mu_n C_{ox}(W/L)}}{2\pi(C_L + C_{gd})}\sqrt{I_D}$$

(b) For $\mu_n C_{ox} = 400\ \mu A/V^2$, $W/L = 20$, $C_L = 20$ fF, $C_{gd} = 5$ fF, and $V_A = 10$ V, provide in table form f_t (GHz), V_{OV} (V), g_m (mA/V), r_o (kΩ), R_o (MΩ), A_M (V/V), and f_H (MHz) for $I_D = 100\ \mu A$, 200 μA, and 500 μA.

10.61 Consider a bipolar cascode amplifier biased at a current of 0.5 mA. The transistors used have $\beta = 100$, $r_o = 100$ kΩ, $C_\pi = 10$ pF, $C_\mu = 2$ pF, and $C_{cs} = 0$. The amplifier is fed with a signal source having $R_{sig} = 5$ kΩ. The load resistance $R_L = 5$ kΩ. Find the low-frequency gain A_M, and estimate the value of the 3-dB frequency f_H.

***10.62** Consider the frequency response of the bipolar cascode amplifier in a situation where r_o can be neglected.

(a) Refer to the circuit in Fig. 10.22, and note that the total resistance between the collector of Q_1 and ground will be equal to r_{e2}, which is usually very small. It follows that the pole introduced at this node will typically be at a very high frequency and thus will have negligible effect on f_H. It also follows that at the frequencies of interest, the gain from the base to the collector of Q_1 will be $-g_{m1}r_{e2} \simeq -1$. Use this to find the capacitance at the

input of Q_1 and hence show that the pole introduced at the input node will have a frequency

$$f_{P1} \simeq \frac{1}{2\pi R'_{sig}(C_{\pi 1} + 2C_{\mu 1})}$$

Then show that the pole introduced at the output node will have a frequency

$$f_{P2} \simeq \frac{1}{2\pi R_L(C_L + C_{cs2} + C_{\mu 2})}$$

(b) Evaluate f_{P1} and f_{P2}, and estimate f_H for the amplifier with $I = 1$ mA, $C_\pi = 10$ pF, $C_\mu = 2$ pF, $C_{cs} = C_L = 0$, $\beta = 100$, and $R_L = R_{sig} = 2$ kΩ.

(c) If R_{sig} is reduced, at what value will the two poles become coincident?

10.63 A BJT cascode amplifier uses transistors for which $\beta = 100$, $V_A = 100$ V, $f_T = 1$ GHz, and $C_\mu = 0.1$ pF. It operates at a bias current of 0.1 mA between a source with $R_{sig} = r_\pi$ and a load $R_L = \beta r_o$. Let $C_L = C_{cs} = 0$. Find the overall voltage gain at dc. By evaluating the various components of τ_H show that the pole introduced at the output node is dominant. Find its frequency and hence an estimate of f_H and the gain–bandwidth product.

Section 10.5: High-Frequency Response of Source and Emitter Followers

10.64 A source follower has $g_m = 5$ mA/V, $g_{mb} = 0$, $r_o = 20$ kΩ, $R_{sig} = 20$ kΩ, $R_L = 2$ kΩ, $C_{gs} = 2$ pF, $C_{gd} = 0.1$ pF, and $C_L = 1$ pF. Find A_M, R_o, f_Z, the frequencies of the two poles, and an estimate of f_H.

10.65 Using the expression for the source follower f_H in Eq. (10.107) show that for situations in which $C_L = 0$, R_{sig} is large, and R_L is small,

$$f_H \simeq \frac{1}{2\pi R_{sig}\left(C_{gd} + \dfrac{C_{gs}}{1 + g_m R'_L}\right)}$$

Find f_H for the case $R_{sig} = 100$ kΩ, $R_L = 2$ kΩ, $r_o = 20$ kΩ, $g_m = 5$ mA/V, $C_{gs} = 12$ pF, and $C_{gd} = 0.1$ pF.

10.66 Refer to Fig. 10.23(c). In situations in which R_{sig} is large, the high-frequency response of the source follower is determined by the low-pass circuit formed by R_{sig} and the input capacitance. An estimate of C_{in} can be obtained by using the Miller approximation to replace C_{gs} with an input capacitance $C_{eq} = C_{gs}(1 - K)$ where K is the gain from gate to source. Using the low-frequency value of $K = g_m R'_L/(1 + g_m R'_L)$ find C_{eq} and hence C_{in} and an estimate of f_H.

10.67 A source follower has a maximally flat gain response with a dc gain of 0.8 and a 3-dB frequency of 1 MHz. Give its transfer function.

10.68 A discrete-circuit source follower driven with $R_{sig} = 100$ kΩ has $C_{gs} = 10$ pF, $C_{gd} = 1$ pF, $C_L = 10$ pF, $g_{mb} = 0$, and r_o very large. The transfer function of the source follower is measured as R_L is varied. At what value of R_L will the transfer function be maximally flat? At this value of R_L the dc gain is found to be 0.9 V/V. What is the 3-dB frequency? What is the value of g_m at which the source follower is operating?

10.69 For an emitter follower biased at $I_C = 1$ mA, having $R_{sig} = R_L = 1$ kΩ, and using a transistor specified to have $f_T = 2$ GHz, $C_\mu = 0.1$ pF, $C_L = 0$, $\beta = 100$, and $V_A = 20$ V, evaluate the low-frequency gain A_M, the frequency of the transmission zero, the pole frequencies, and an estimate of the 3-dB frequency f_H.

Section 10.6: High-Frequency Response of Differential Amplifiers

10.70 A MOSFET differential amplifier such as that shown in Fig. 10.26(a) is biased with a current source $I = 200$ μA. The transistors have $W/L = 12.5$, $k'_n = 400$ μA/V^2, $V_A = 5$ V, $C_{gs} = 20$ fF, $C_{gd} = 5$ fF, and $C_{db} = 5$ fF. The drain resistors are 50 kΩ each. Also, there is a 20-fF capacitive load between each drain and ground.

(a) Find V_{OV} and g_m for each transistor.

(b) Find the differential gain A_d.

(c) If the input signal source has a small resistance R_{sig} and thus the frequency response is determined primarily by the output pole, estimate the 3-dB frequency f_H.

(d) If, in a different situation, the amplifier is fed symmetrically with a signal source of 40 kΩ resistance (i.e., 20 kΩ in series with each gate terminal), use the open-circuit time-constants method to estimate f_H.

10.71 A MOS differential amplifier is biased with a current source having an output resistance $R_{SS} = 200$ kΩ and an output capacitance $C_{SS} = 0.25$ pF. If the differential gain is found to have a dominant pole at 20 MHz, what is the 3-dB frequency of the CMRR?

10.72 The differential gain of a MOS amplifier is 1000 V/V with a dominant pole at 20 MHz. The common-mode gain is 0.1 V/V at low frequencies and has a transmission zero at 200 kHz. Sketch a Bode plot for the CMRR.

10.73 A differential amplifier is biased by a current source having an output resistance of 100 kΩ and an output

capacitance of 100 fF. The differential gain exhibits a dominant pole at 40 MHz. What are the poles of the CMRR?

10.74 In a particular MOS differential amplifier design, the bias current $I = 100\ \mu A$ is provided by a single transistor operating at $V_{OV} = 0.4$ V with $V_A = 40$ V and output capacitance C_{SS} of 100 fF. What is the frequency of the common-mode gain zero (f_Z) at which A_{cm} begins to rise above its low-frequency value? To meet a requirement for reduced power supply, you consider reducing V_{OV} to 0.2 V while keeping I unchanged. Assuming the current-source capacitance to be directly proportional to the device width, what is the impact on f_Z of this proposed change?

10.75 A BJT differential amplifier operating with a 0.5-mA current source uses transistors for which $\beta = 100$, $f_T = 600$ MHz, and $C_\mu = 0.4$ pF. Each of the collector resistances is 10 kΩ, and r_o is very large. The amplifier is fed in a symmetrical fashion with a source resistance of 10 kΩ in series with each of the two input terminals.

(a) Sketch the differential half-circuit and its high-frequency equivalent circuit.
(b) Determine the low-frequency value of the overall differential gain.
(c) Use the Miller effect to determine the input capacitance and hence estimate the 3-dB frequency f_H and the gain–bandwidth product.

10.76 A current-mirror-loaded MOS differential amplifier is biased with a current source $I = 0.2$ mA. The two NMOS transistors of the differential pair are operating at $V_{OV} = 0.2$ V, and the PMOS devices of the mirror are operating at $|V_{OV}| = 0.2$ V. The Early voltage $V_{An} = |V_{Ap}| = 5$ V. The total capacitance at the input node of the mirror is 0.05 pF and that at the output node of the amplifier is 0.1 pF. Find the dc value and the frequencies of the poles and zero of the differential voltage gain.

***10.77** Consider the current-mirror-loaded CMOS differential amplifier of Fig. 10.29(a) for the case of all transistors operated at the same $|V_{OV}|$ and having the same $|V_A|$. Also let the total capacitance at the output node (C_L) be four times the total capacitance at the input node of the current mirror C_m. Give expressions for A_d, f_{P1}, f_{P2}, and f_Z. Hence show that $f_{P2}/f_{P1} = 4A_d$ and $f_t = g_m/2\pi C_L$. For $V_A = 20$ V, $V_{OV} = 0.2$ V, $I = 0.2$ mA, $C_L = 100$ fF, and $C_m = 25$ fF, find the dc value of A_d, and the value of f_{P1}, f_t, f_{P2}, and f_Z and sketch a Bode plot for $|A_d|$.

***10.78** For the current mirror in Fig. P10.78, derive an expression for the current transfer function $I_o(s)/I_i(s)$ taking into account the BJT internal capacitances and neglecting r_o. Assume the BJTs to be identical. Observe that a signal ground appears at the collector of Q_2. If the mirror is biased at 1 mA and the BJTs at this operating point are characterized by $f_T = 500$ MHz, $C_\mu = 2$ pF, and $\beta_0 = 100$, find the frequencies of the pole and zero of the transfer function.

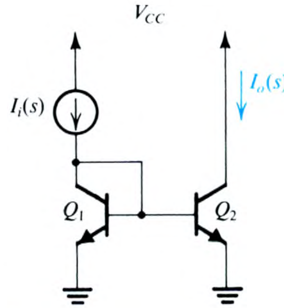


Figure P10.78

Section 10.7: Other Wideband Amplifier Configurations

10.79 A CS amplifier is specified to have $g_m = 4$ mA/V, $r_o = 40$ kΩ, $C_{gs} = 30$ fF, $C_{gd} = 5$ fF, $C_L = 20$ fF, $R_{sig} = 40$ kΩ, and $R_L = 40$ kΩ.

(a) Find the low-frequency gain A_M, and use open-circuit time constants to estimate the 3-dB frequency f_H. Hence determine the gain–bandwidth product.
(b) If a 500-Ω resistance is connected in the source lead, find the new values of $|A_M|$, f_H, and the gain–bandwidth product.

10.80 In a discrete-circuit CS amplifier a source-degeneration resistance is being used to control the bandwidth. Assume that r_o is very large and C_L is negligibly small. Adapt the formulas given in the text for this case and thus give the expressions for A_M and f_H. Let $R_{sig} = 100$ kΩ, $g_m = 5$ mA/V, $R_L = 5$ kΩ, $C_{gs} = 10$ pF, and $C_{gd} = 2$ pF. Find $|A_M|$, f_H, and the gain–bandwidth product for these three cases: $R_s = 0$, 100 Ω, and 200 Ω.

D 10.81 (a) Use the approximate expression in Eq. (10.139) to determine the gain–bandwidth product of a CS amplifier with a source-degeneration resistance. Assume $C_{gd} = 0.1$ pF and $R_{sig} = 500$ kΩ.
(b) If a low-frequency gain of 20 V/V is required, what f_H corresponds?
(c) For $g_m = 4$ mA/V, $A_0 = 100$ V/V, and $R_L = 40$ kΩ, find the required value of R_s.

10.82 For the CS amplifier with a source-degeneration resistance R_s, show for $R_{sig} \gg R_s$, $r_o \gg R_s$, and $R_L = r_o$ that

$$A_M = \frac{-A_0}{2+k}$$

and

$$\tau_H \simeq \frac{C_{gs}R_{sig}}{1+(k/2)} + C_{gd}R_{sig}\left(1+\frac{A_0}{2+k}\right) + (C_L + C_{gd})r_o\left(\frac{1+k}{2+k}\right)$$

where $k \equiv g_m R_s$

D *10.83 Generate a table of $|A_M|$, f_H, and GB versus $k \equiv g_m R_s$ for a CS amplifier with a source-degeneration resistance R_s. The table should have entries for $k = 0, 1, 2, \ldots, 15$. The amplifier is specified to have $g_m = 5$ mA/V, $r_o = 40$ kΩ, $R_L = 40$ kΩ, $R_{sig} = 20$ kΩ, $C_{gs} = 2$ pF, $C_{gd} = 0.1$ pF, and $C_L = 1$ pF. Use the formulas for A_M and τ_H given in the statement for Problem 10.82. If $f_H = 2$ MHz is required, find the value needed for R_s and the corresponding value of $|A_M|$.

***10.84** In this problem we investigate the bandwidth extension obtained by placing a source follower between the signal source and the input of the CS amplifier.

(a) First consider the CS amplifier of Fig. P10.84(a). Show that

$$A_M = -g_m r_o$$
$$\tau_H = C_{gs}R_{sig} + C_{gd}\left[R_{sig}(1+g_m r_o) + r_o\right] + C_L r_o$$

where C_L is the total capacitance between the output node and ground. Calculate the value of A_M, f_H, and the gain–bandwidth product for the case $g_m = 1$ mA/V, $r_o = 20$ kΩ, $R_{sig} = 20$ kΩ, $C_{gs} = 20$ fF, $C_{gd} = 5$ fF, and $C_L = 10$ fF.

(b) For the CD–CS amplifier in Fig. P10.84(b), show that

$$A_M = -\frac{r_{o1}}{1/g_{m1} + r_{o1}}(g_{m2}r_{o2})$$
$$\tau_H = C_{gd1}R_{sig} + C_{gs1}\frac{R_{sig}+r_{o1}}{1+g_{m1}r_{o1}} + C_{gs2}\left(\frac{1}{g_{m1}} \parallel r_{o1}\right) + C_{gd2}\left[\left(\frac{1}{g_{m1}} \parallel r_{o1}\right)(1+g_{m2}r_{o2}) + r_{o2}\right] + C_L r_{o2}$$

Calculate the values of A_M, f_H, and the gain–bandwidth product for the same parameter values used in (a). Compare with the results of (a).

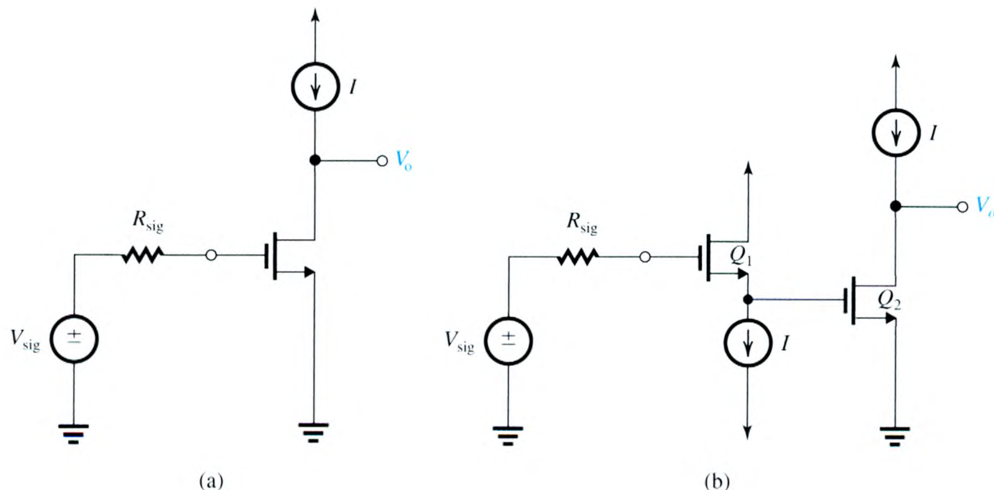


Figure P10.84

***10.85** The transistors in the circuit of Fig. P10.85 have $\beta_0 = 100$, $V_A = 100$ V, and $C_\mu = 0.2$ pF. At a bias current of 100 μA, $f_T = 200$ MHz. (Note that the bias details are not shown.)

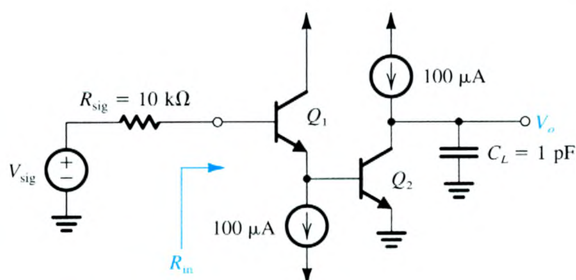


Figure P10.85

(a) Find R_{in} and the midband gain.
(b) Find an estimate of the upper 3-dB frequency f_H. Which capacitor dominates? Which one is the second most significant?

(*Hint*: Use the formulas in Example 10.10.)

10.86 Consider the circuit of Fig. P10.86 for the case: $I = 200$ μA and $V_{OV} = 0.2$ V, $R_{sig} = 100$ kΩ, $R_D = 40$ kΩ, $C_{gs} = 2$ pF, and $C_{gd} = 0.2$ pF. Find the dc gain, the high-frequency poles, and an estimate of f_H. Neglect r_o.

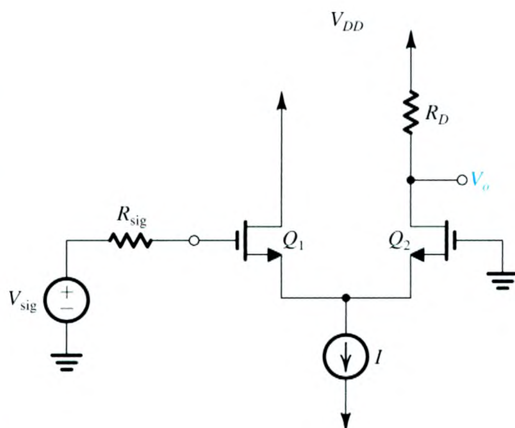


Figure P10.86

10.87 For the amplifier in Fig. 10.34(a), let $I = 1$ mA, $\beta = 120$, $f_T = 500$ MHz, and $C_\mu = 0.5$ pF, and neglect r_o. Assume that a load resistance of 10 kΩ is connected to the output terminal. If the amplifier is fed with a signal V_{sig} having a source resistance $R_{sig} = 12$ kΩ, find A_M and f_H.

10.88 Consider the CD–CG amplifier of Fig. 10.34(c) for the case $g_m = 5$ mA/V, $C_{gs} = 2$ pF, $C_{gd} = 0.1$ pF, C_L (at the output node) = 1 pF, and $R_{sig} = R_L = 20$ kΩ. Neglecting r_o, find A_M and f_H. (*Hint:* Evaluate f_H directly from the transfer function.)

***10.89** Figure P10.89 shows an amplifier formed by cascading two CS stages. Note that the input bias voltage is not shown. Each of Q_1 and Q_2 is operated at an overdrive voltage of 0.2 V, and $|V_A| = 10$ V. The transistor capacitances are as follows: $C_{gs} = 20$ fF, $C_{gd} = 5$ fF, and $C_{db} = 5$ fF. The signal-source resistance $R_{sig} = 10$ kΩ. Also, the output resistance of each of the current sources is equal to the MOSFET r_o.

(a) Find the dc voltage gain.
(b) Use the method of open-circuit time constants to determine an estimate for the 3-dB frequency f_H.

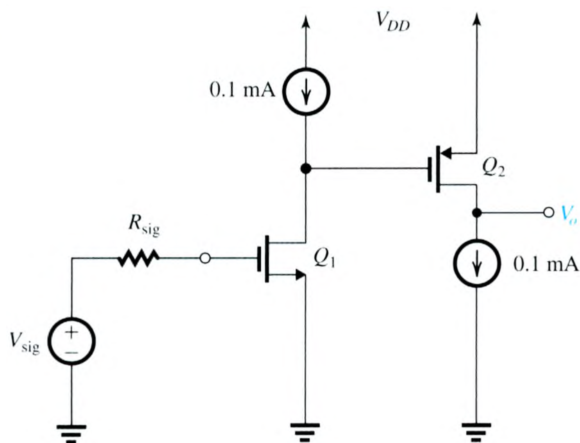


Figure P10.89

****10.90** Consider the BiCMOS amplifier shown in Fig. P10.90. The BJT has $|V_{BE}| = 0.7$ V, $\beta = 200$,

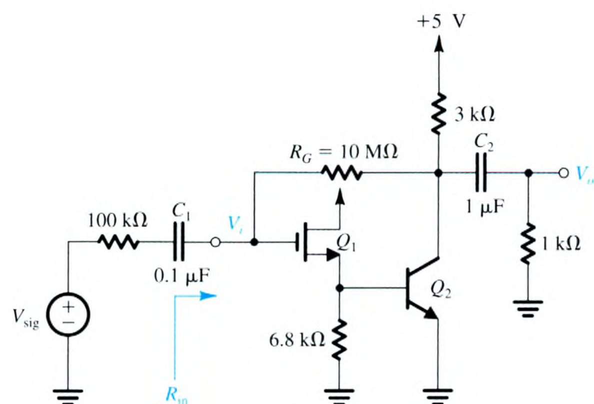


Figure P10.90

$C_\mu = 0.8$ pF, and $f_T = 600$ MHz. The NMOS transistor has $V_t = 1$ V, $k_n' W/L = 2$ mA/V^2, and $C_{gs} = C_{gd} = 1$ pF.

(a) Consider the dc bias circuit. Neglect the base current of Q_2 in determining the current in Q_1. Find the dc bias currents in Q_1 and Q_2, and show that they are approximately 100 µA and 1 mA, respectively.

(b) Evaluate the small-signal parameters of Q_1 and Q_2 at their bias points.

(c) Consider the circuit at midband frequencies. First, determine the small-signal voltage gain V_o/V_i. (Note that R_G can be neglected in this process.) Then use Miller's theorem on R_G to determine the amplifier input resistance R_{in}. Finally, determine the overall voltage gain V_o/V_{sig}. Assume r_o of both transistors to be very large.

(d) Consider the circuit at higher frequencies. Use Miller's theorem to replace R_G with a resistance at the input. (The one at the output will be too large to matter.) Use open-circuit time constants to estimate f_H.

*****10.91** In each of the six circuits in Fig. P10.91, let $\beta = 100$, $C_\mu = 2$ pF, and $f_T = 400$ MHz, and neglect r_o. Calculate the midband gain A_M, the 3-dB frequency f_H, and the gain–bandwidth product. Provide a summary of your results in a table with the following columns: Case, Configuration Name, A_M (V/V), f_H (MHz), and GB (MHz).

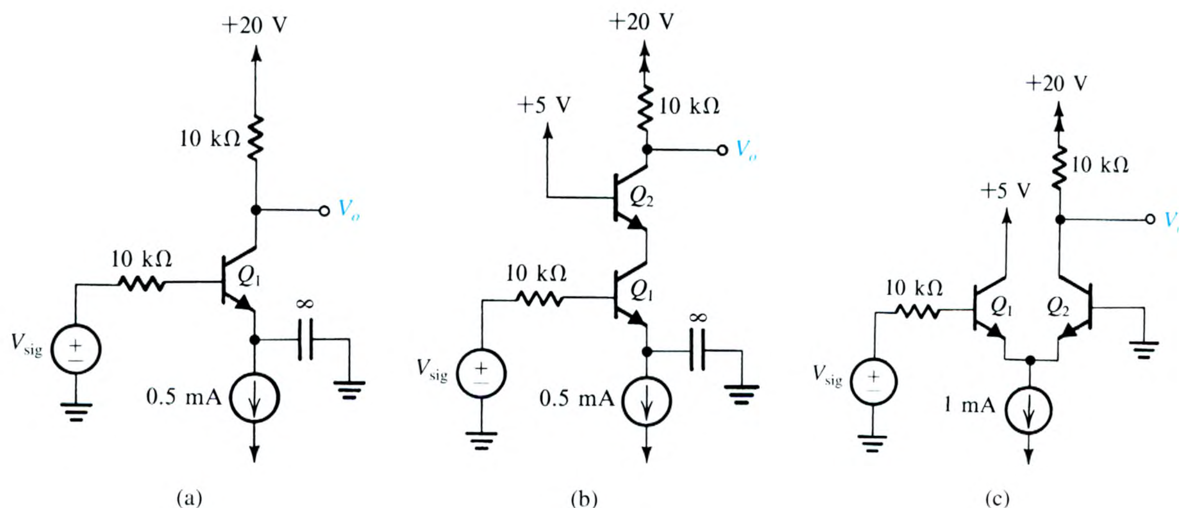


Figure P10.91

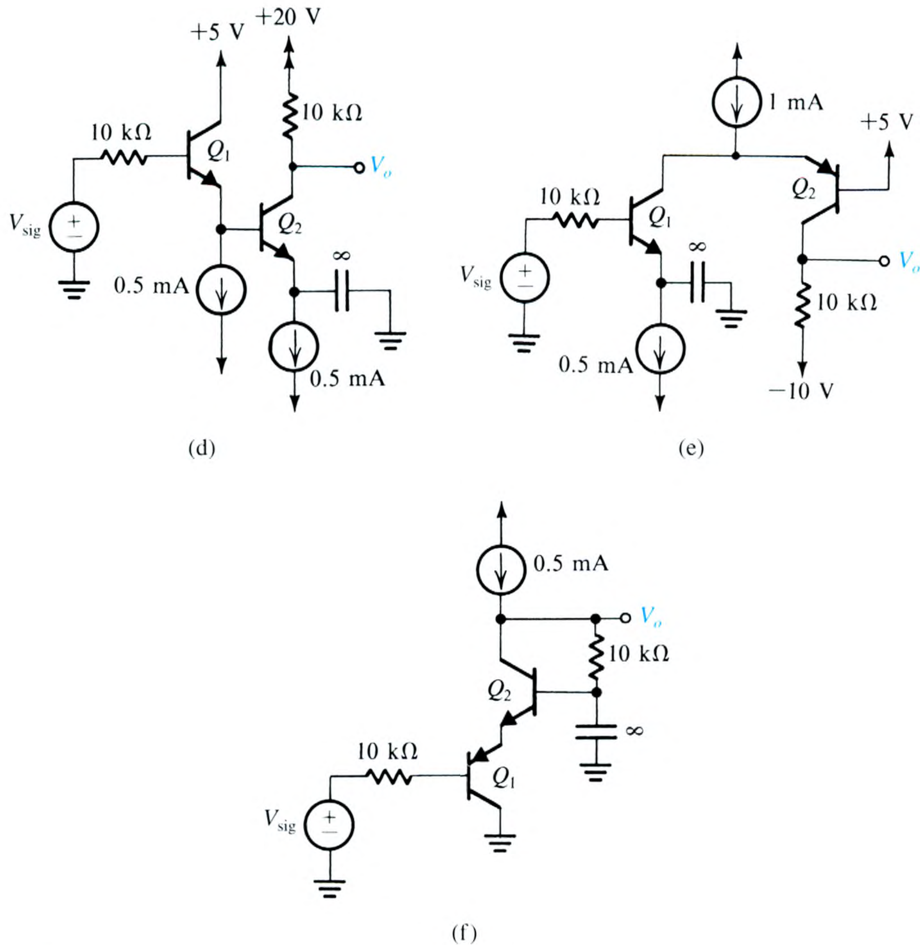


Figure P10.91 *continued*

Section 10.8: Low-Frequency Response of Discrete-Circuit CS and CE Amplifiers

10.92 A signal source V_{sig} with a resistance $R_{sig} = 1\ k\Omega$ is connected to a load resistance $R_L = 4\ k\Omega$ via a capacitor $C = 1\ \mu F$. Find:

(a) the transfer function V_o/V_{sig};
(b) the magnitude of transmission at high frequencies for which the coupling capacitor acts as a short circuit;
(c) the frequency at which $|V_o/V_{sig}|$ is 3 dB below the high-frequency value; and
(d) the transmission at dc.

D 10.93 For the amplifier in Fig. 10.37(a), if $R_{G1} = 3\ M\Omega$, $R_{G2} = 2\ M\Omega$, and $R_{sig} = 500\ k\Omega$, find the value of the coupling capacitor C_{C1} (specified to one significant digit) that places the associated pole at 10 Hz or lower.

D 10.94 For the amplifier in Fig. 10.37(a), if $R_D = 20\ k\Omega$, $R_L = 20\ k\Omega$, and r_o is very large, find the value of C_{C2} (specified to one significant digit) that places the associated pole at 10 Hz or lower.

D 10.95 The amplifier in Fig. 10.37(a) is biased to operate at $g_m = 4$ mA/V, and $R_S = 2\ k\Omega$. Find the value of C_S (specified to one significant digit) that places its associated

pole at 100 Hz or lower. What are the actual frequencies of the pole and zero realized?

10.96 The amplifier in Fig. 10.37(a) is biased to operate at $g_m = 5$ mA/V, and has the following component values: $R_{sig} = 100$ kΩ, $R_{G1} = 47$ MΩ, $R_{G2} = 10$ MΩ, $C_{C1} = 0.01$ μF, $R_S = 2$ kΩ, $C_S = 10$ μF, $R_D = 4.7$ kΩ, $R_L = 10$ kΩ, and $C_{C2} = 1$ μF. Find A_M, f_{P1}, f_{P2}, f_Z, f_{P3}, and f_L.

D 10.97 The amplifier in Fig. P10.97 is biased to operate at $g_m = 2$ mA/V. Neglect r_o.

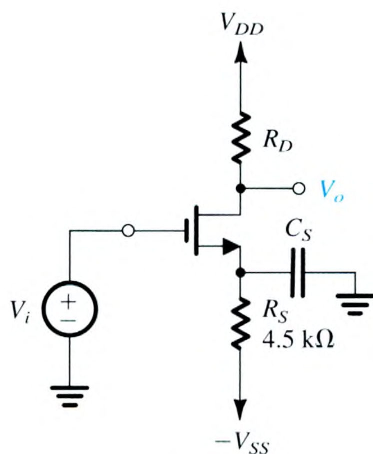


Figure P10.97

(a) Determine the value of R_D that results in a midband gain of −20 V/V.

(b) Determine the value of C_S that results in a pole frequency of 100 Hz.

(c) What is the frequency of the transmission zero introduced by C_S?

(d) Give an approximate value for the 3-dB frequency f_L.

(e) Sketch a Bode plot for the gain of this amplifier. What does the plot tell you about the gain at dc? Does this make sense? Why or why not?

D 10.98 Figure P10.98 shows a CS amplifier biased by a constant-current source I. Let $R_{sig} = 0.4$ MΩ, $R_G = 1.5$ MΩ, $g_m = 4$ mA/V, $R_D = 20$ kΩ, and $R_L = 20$ kΩ. Find A_M. Also, design the coupling and bypass capacitors to locate the three low-frequency poles at 100 Hz, 10 Hz, and 1 Hz. Design so as to minimize the total capacitance, with the capacitors specified only to a single significant digit. What value of f_L results?

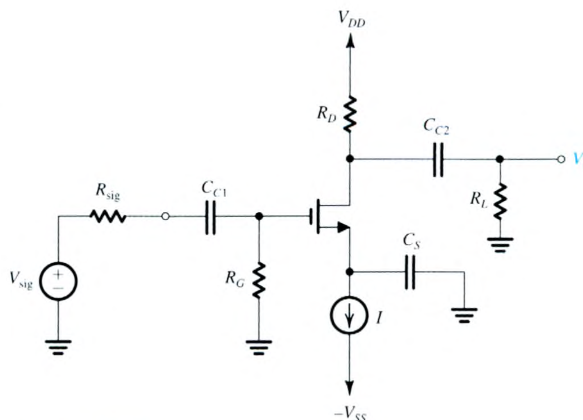


Figure P10.98

D 10.99 Figure P10.99 shows a current-source-biased CE amplifier operating at 100 μA from ±3-V power supplies. It employs $R_C = 20$ kΩ, $R_B = 200$ kΩ, and operates between a 20-kΩ source and a 10-kΩ load. The transistor $\beta = 100$. Select C_E first, for a minimum value specified to one significant digit and providing up to 80% of f_L where f_L is to be 100 Hz. Then choose C_{C1} and C_{C2}, each specified to one significant digit, and each contributing about 10% of f_L. What f_L results? What total capacitance is needed?

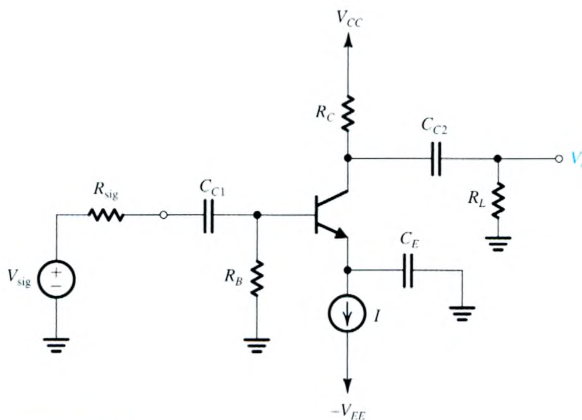


Figure P10.99

10.100 Consider the common-emitter amplifier of Fig. 10.41(a) under the following conditions: $R_{sig} = 5$ kΩ, $R_{B1} = 33$ kΩ, $R_{B2} = 22$ kΩ, $R_E = 3.9$ kΩ, $R_C = 4.7$ kΩ, $R_L = 5.6$ kΩ, $V_{CC} = 5$ V. The dc emitter current can be shown to be $I_E \simeq 0.3$ mA, at which $\beta = 120$. Find the input resistance R_{in} and the midband gain A_M. If $C_{C1} = C_{C2} = 1$ μF and $C_E = 20$ μF, find the three short-circuit time constants and an estimate for f_L.

D 10.101 For the amplifier described in Problem 10.100, design the coupling and bypass capacitors for a lower 3-dB frequency of 200 Hz. Design so that the contribution of each of C_{C1} and C_{C2} to determining f_L is only 10%.

10.102 Consider the circuit of Fig. 10.41(a). For $R_{sig} = 5\text{ k}\Omega$, $R_B \equiv R_{B1} \| R_{B2} = 10\text{ k}\Omega$, $r_\pi = 1\text{ k}\Omega$, $\beta_0 = 100$, and $R_E = 1.5\text{ k}\Omega$, what is the ratio C_E/C_{C1} that makes their contributions to the determination of f_L equal?

D *10.103 For the common-emitter amplifier of Fig. P10.103, neglect r_o and assume the current source to be ideal.

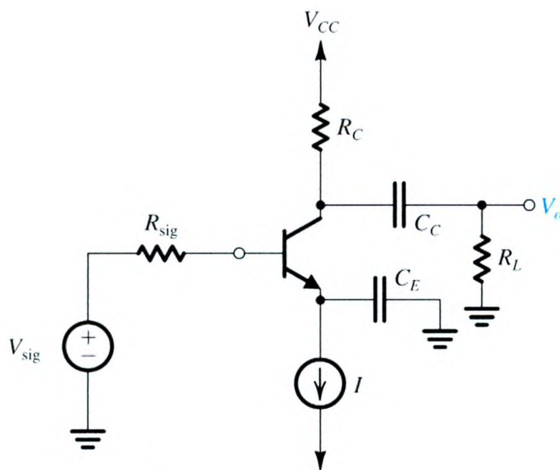


Figure P10.103

(a) Derive an expression for the midband gain.

(b) Convince yourself that the two poles caused by C_E and C_C do not interact. Find expressions for their frequencies, ω_{PE} and ω_{PC}.

(c) Give an expression for the amplifier voltage gain $V_o(s)/V_{sig}(s)$ in terms of A_M, ω_{PE}, and ω_{PC}.

(d) For $R_{sig} = R_C = R_L = 10\text{ k}\Omega$, $\beta = 100$, and $I = 1\text{ mA}$, find the value of the midband gain.

(e) Select values for C_E and C_C to place the two pole frequencies a decade apart and to obtain a lower 3-dB frequency of 100 Hz while minimizing the total capacitance.

(f) Sketch a Bode plot for the gain magnitude, and estimate the frequency at which the gain becomes unity.

***10.104** The BJT common-emitter amplifier of Fig. P10.104 includes an emitter-degeneration resistance R_e.

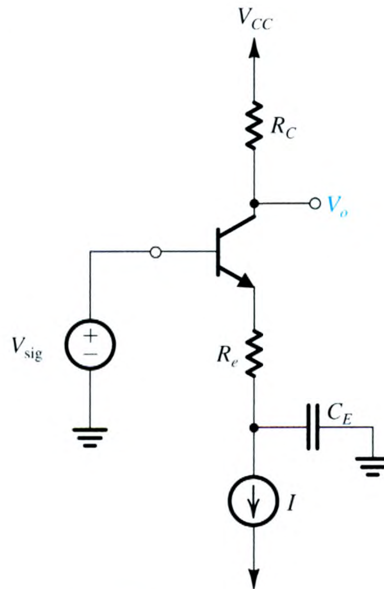


Figure P10.104

(a) Assuming $\alpha \simeq 1$, neglecting r_o, and assuming the current source to be ideal, derive an expression for the small-signal voltage gain $A(s) \equiv V_o/V_{sig}$ that applies in the midband and the low-frequency band. Hence find the midband gain A_M and the lower 3-dB frequency f_L.

(b) Show that including R_e reduces the magnitude of A_M by a certain factor. What is this factor?

(c) Show that including R_e reduces f_L by the same factor as in (b) and thus one can use R_e to trade off gain for bandwidth.

(d) For $I = 0.25\text{ mA}$, $R_C = 10\text{ k}\Omega$, and $C_E = 10\ \mu\text{F}$, find $|A_M|$ and f_L with $R_e = 0$. Now find the value of R_e that lowers f_L by a factor of 10. What will the gain become? Sketch on the same diagram a Bode plot for the gain magnitude for both cases.

CHAPTER 11

Feedback

IN THIS CHAPTER YOU WILL LEARN

- The general structure of the negative-feedback amplifier and the basic principle that underlies its operation.
- The advantages of negative feedback, how these come about, and at what cost.
- The appropriate feedback topology to use with amplifiers of each of the four types: voltage, current, transconductance, and transresistance.
- An intuitive, insightful approach for analyzing practical feedback-amplifier circuits.
- Why and how negative-feedback amplifiers can become unstable (i.e., oscillate) and how to design the circuit to ensure stable performance.

Introduction

Most physical systems incorporate some form of feedback. It is interesting to note, though, that the theory of negative feedback has been developed by electronics engineers. In his search for methods for designing amplifiers with stable gain for use in transatlantic telephone repeaters, Harold Black, an electronics engineer with the Western Electric Company, invented the feedback amplifier in 1928. Since then, the technique has been so widely used that it is almost impossible to think of electronic circuits without some form of feedback, either implicit or explicit. Furthermore, feedback and its associated theory are currently used in areas other than engineering, such as in modeling biological systems.

Feedback can be either **negative** or **positive**. In amplifier design, negative feedback is applied to achieve one or more of the following goals:

1. *Desensitize the gain:* that is, make the value of the gain less sensitive to variations in the values of circuit components, such as might be caused by changes in temperature.
2. *Reduce nonlinear distortion:* that is, make the output proportional to the input (in other words, make the gain constant, independent of signal level).
3. *Reduce the effect of noise:* that is, minimize the contribution to the output of unwanted electric signals generated, either by the circuit components themselves or by extraneous interference.
4. *Control the input and output resistances:* that is, raise or lower the input and output resistances by the selection of an appropriate feedback topology.
5. *Extend the bandwidth* of the amplifier.

All of the desirable properties above come at the expense of a reduction in gain. It will be shown that the gain-reduction factor, called the **amount of feedback**, is the factor by which the circuit is desensitized, by which the input resistance of a voltage amplifier is increased, by which the bandwidth is extended, and so on. In short, *the basic idea of negative feedback is to trade off gain for other desirable properties*. This chapter is devoted to the analysis and design of negative-feedback amplifiers.

Under certain conditions, the negative feedback can introduce undesirable behavior in the dynamic response of the amplifier. This can eventually lead to oscillation, that is, the generation of unwanted signals. In such a case, the negative feedback has turned positive. In fact, in Chapter 15 we will study the use of positive feedback in the design of oscillators and bistable circuits. Here, however, we are interested in the design of stable amplifiers. We shall therefore study the stability problem of negative-feedback amplifiers and their potential for oscillation.

Positive feedback does not always lead to instability. In fact, positive feedback is useful in a number of nonregenerative applications. An example is the design of active filters, which are studied in Chapter 14.

Before we begin our study of negative feedback, recall that we have already encountered negative feedback in a number of applications: Almost all op-amp circuits (Chapter 2) employ negative feedback. Another popular application of negative feedback is the use of the emitter resistance R_E to stabilize the bias point of bipolar transistors and to increase the input resistance, bandwidth, and linearity of a BJT amplifier. In addition, the source follower and the emitter follower both use a large amount of negative feedback. You may wonder

now about the need for a formal study of negative feedback, but by the end of this chapter, you should appreciate that an understanding of feedback provides an invaluable tool for the analysis and design of electronic circuits. Also, the insight gained by thinking in terms of feedback can be extremely profitable.

11.1 The General Feedback Structure

11.1.1 Signal-Flow Diagram

Figure 11.1 shows the basic structure of a feedback amplifier. Rather than showing voltages and currents, Fig. 11.1 is a signal-flow diagram, where each of the quantities x can represent either a voltage or a current signal. The basic amplifier is *unilateral* and has a gain A, known as the **open-loop gain**; thus its output x_o is related to the input x_i by

$$x_o = Ax_i \tag{11.1}$$

The *feedback network* measures or samples the output signal x_o and provides a *feedback signal* x_f that is related to x_o by the **feedback factor**[1] β,

$$x_f = \beta x_o \tag{11.2}$$

We are assuming that connecting the feedback network to the amplifier output does not change the gain A or the value of x_o; that is, *the feedback network does not load the amplifier output*. Also, the feedback network is unilateral.

The feedback signal x_f is *subtracted* from the source signal x_s, which is the input to the complete feedback amplifier,[2] to produce the signal x_i, which is the input to the basic amplifier,

$$x_i = x_s - x_f \tag{11.3}$$

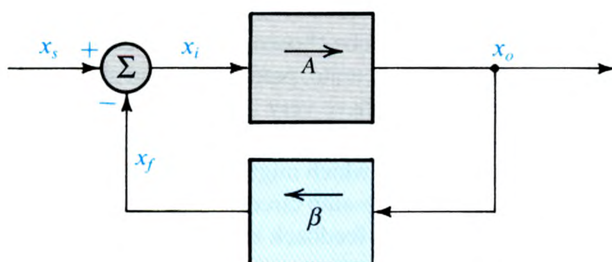


Figure 11.1 General structure of the feedback amplifier. This is a signal-flow diagram, and the quantities x represent either voltage or current signals.

[1]The feedback factor β has nothing to do with the current gain β of the common-emitter transistor. To avoid possible confusion, we will sometimes use h_{fe} to denote the current gain of the BJT. The symbol h_{fe} comes from the hybrid or h-parameters of two-port networks (See Appendix C).

[2]In earlier chapters, we used the subscript "sig" for quantities associated with the signal source (e.g., v_{sig} and R_{sig}). We did that to avoid confusion with the subscript "s," which is usually used with FETs to denote quantities associated with the source terminal of the transistor. At this point, however, it is expected that readers have become sufficiently familiar with the subject that the possibility of confusion is minimal. Therefore, we will revert to using the simpler subscript s for signal-source quantities.

It is this subtraction that makes the feedback negative. In essence, *negative feedback reduces the signal that appears at the input of the basic amplifier.* Here, too, we assume that connecting the output of the feedback network to the amplifier input, through the subtractor or differencing circuit, does not change the gain A; that is, *the feedback network does not load the amplifier input.*

11.1.2 The Closed-Loop Gain

The gain of the feedback amplifier, known as the closed-loop gain or the **gain-with-feedback** and denoted A_f, is defined as

$$A_f \equiv \frac{x_o}{x_s}$$

Combining Eqs. (11.1) through (11.3) provides the following expression for A_f:

$$A_f = \frac{A}{1+A\beta} \tag{11.4}$$

The quantity $A\beta$ is called the **loop gain**, a name that follows from Fig. 11.1. For the feedback to be negative, the loop gain $A\beta$ must be positive; that is, the feedback signal x_f should have the same sign as x_s, resulting in a smaller difference signal x_i. Equation (11.4) indicates that for positive $A\beta$ the gain with feedback A_f will be smaller than the open-loop gain A by a factor equal to $1 + A\beta$, which is called the **amount of feedback**.

If, as is the case in many circuits, the loop gain $A\beta$ is large, $A\beta \gg 1$, then from Eq. (11.4) it follows that

$$A_f \simeq \frac{1}{\beta} \tag{11.5}$$

which is a very interesting result: *When the loop gain is large, the gain of the feedback amplifier is almost entirely determined by the feedback network.* Since the feedback network usually consists of passive components, which usually can be chosen to be as accurate as we want, the advantage of negative feedback in obtaining accurate, predictable, and stable gain should be apparent. In other words, the overall gain will have very little dependence on the gain of the basic amplifier, A, a desirable property because the gain A is usually a function of many manufacturing and application parameters, some of which might have wide tolerances. We saw a dramatic illustration of all of these effects in op-amp circuits in Chapter 2, where the closed-loop gain is almost entirely determined by the feedback elements. Generally, we will consider $(1/\beta)$ to be the ideal value of A_f,

$$A_f\,|_{\text{ideal}} = \frac{1}{\beta} \tag{11.6}$$

The deviation of A_f from the ideal value can be quantified by writing the expression in Eq. (11.4) in the form

$$A_f = \left(\frac{1}{\beta}\right)\frac{A\beta}{1+A\beta} \tag{11.7}$$

Thus,

$$A_f = A_f\,|_{\text{ideal}}\ \frac{1}{1+(1/A\beta)} \tag{11.8}$$

As an example, a loop gain $A\beta = 100$ leads to a closed-loop gain that is 1% below its ideal value of $(1/\beta)$.

Equations (11.1) through (11.3) can be combined to obtain the following expression for the feedback signal x_f:

$$x_f = \frac{A\beta}{1+A\beta} x_s \tag{11.9}$$

Thus for $A\beta \gg 1$ we see that $x_f \simeq x_s$, which implies that the signal x_i at the input of the basic amplifier is reduced to almost zero. If a large amount of negative feedback is used, the feedback signal x_f becomes almost a replica of the input signal x_s. The difference between x_s and x_f, which is x_i, is sometimes referred to as the **error signal**.[3] Accordingly, the **input differencing circuit** is often also called a **comparison circuit**. (It is also known as a **mixer**.) An expression for x_i can be easily determined as

$$x_i = \frac{1}{1+A\beta} x_s \tag{11.10}$$

from which we can verify that for $A\beta \gg 1$, x_i becomes very small. Notice that negative feedback reduces the signal that appears at the input terminals of the basic amplifier by the amount of feedback $(1 + A\beta)$. As we will see, it is this reduction of input signal that increases the linearity of the feedback amplifier.[4]

11.1.3 The Loop Gain

From the discussion above we see that the loop gain $A\beta$ is a very important—in fact, the most important—characteristic parameter of a feedback amplifier:

1. The sign of $A\beta$ determines the polarity of the feedback; the loop gain $A\beta$ must be positive for the feedback to be negative.
2. The magnitude of $A\beta$ determines how close the closed-loop gain A_f is to the ideal value of $1/\beta$.
3. The magnitude of $A\beta$ determines the amount of feedback $(1 + A\beta)$ and hence, as we will see in the next section, the magnitude of the various improvements in amplifier performance resulting from the negative feedback.
4. As we will also see in later sections, the inevitable variation of $A\beta$ with frequency can cause the dynamic closed-loop response to exhibit undesirable properties. Also, if the phase shift of $A\beta$ reaches 180°, the feedback becomes positive, which in turn can cause the feedback amplifier to become unstable. It follows that the design of a stable feedback amplifier may involve modifying the frequency behaviors of its loop gain $A\beta$ appropriately (Section 11.10).

Because the loop gain is so important, we should take a moment to consider how to find it. Figure 11.1 indicates that the value of the loop gain $A\beta$ can be determined as follows:

---

[3]This terminology is more common in feedback control systems than in feedback amplifiers.

[4]We have in fact already seen examples of this: adding a resistance R_e in the emitter of a CE amplifier (or a resistance R_s in the source of a CS amplifier) increases the linearity of these amplifiers because for the same input signal as before, v_{be} and v_{gs} are now smaller (by the amount of feedback).

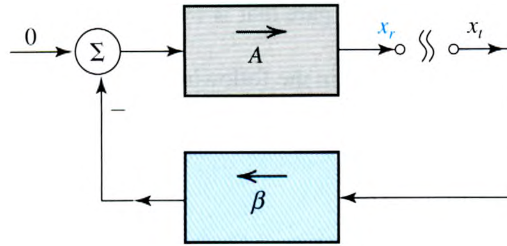


Figure 11.2 Determining the loop gain by breaking the feedback loop at the output of the basic amplifier, applying a test signal x_t, and measuring the returned signal x_r : $A\beta \equiv -x_r/x_t$.

1. Set $x_s = 0$.
2. Break the feedback loop at a convenient location, ensuring that the values of A and *β do not change.* Since we assumed that the feedback network does not load the amplifier output, we can break the loop at the amplifier output (see Fig. 11.2) without causing A to change.
3. Apply a test signal x_t to the input of the loop where the break has been made and determine the *returned signal* x_r at the loop output, that is, at the other side of the break. From Fig. 11.2 we see that

$$x_r = -A\beta x_t$$

and the loop gain $A\beta$ is obtained as

$$A\beta = -\frac{x_r}{x_t} \tag{11.11}$$

Notice that since $A\beta$ is positive, the returned signal x_r will be out of phase with the test signal x_t, verifying that the feedback is indeed negative. In fact, we can use this approach qualitatively to ascertain the polarity of the feedback. We will have a lot more to say about the loop gain in subsequent sections.

11.1.4 The Ideal Case of Infinite Open-Loop Gain *A*

An important limit case is obtained when $A = \infty$ leading to $A\beta = \infty$, and from Eq. (11.8), the closed-loop gain A_f achieves its ideal value of $(1/\beta)$. In this case, Eqs. (11.9) and (11.10) give

$$x_f = x_s \tag{11.12}$$

$$x_i = 0 \tag{11.13}$$

Notice that although $x_i = 0$, the infinite A makes x_o finite and related to x_s by the ideal value of the closed-loop gain,

$$x_o = \left(\frac{1}{\beta}\right)x_s$$

Figure 11.3 shows the signal-flow graph of the feedback amplifier for the case $A = \infty$.

This ideal case should remind you of the analysis of op-amp circuits assuming an ideal op amp with infinite open-loop gain A (Chapter 2). The virtual short circuit that results between the two op-amp input terminals is a manifestation of Eq. (11.13); namely, that the input signal $x_i = 0$.

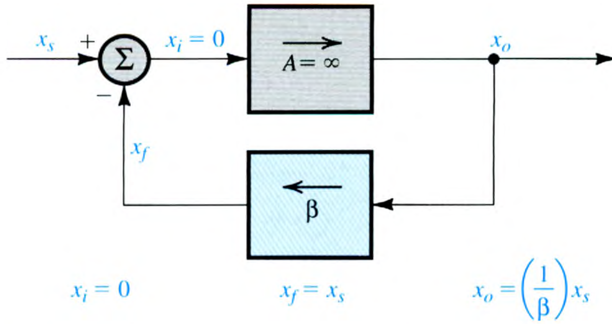


Figure 11.3 Signal-flow graph of a feedback amplifier in the limit case $A = \infty$.

This limit case, although an idealization, is very useful in both the design and analysis of feedback amplifiers: In design, where we want to obtain a given ideal closed-loop gain, we use Eq. (11.6) to determine the required β,

$$\beta = 1/A_f\,|_{\text{ideal}} \tag{11.14}$$

In analysis of a given feedback amplifier circuit, we start by setting the input signal to the open-loop amplifier to zero, and then finding x_o that results for the applied x_s (see Fig. 11.3). The ideal value of the closed-loop gain can then be found as

$$A_f\,|_{\text{ideal}} = \frac{x_o}{x_s}$$

and the feedback factor β is determined using Eq. (11.14). We will illustrate this procedure with many examples in the sections to follow.

Example 11.1

The noninverting op-amp configuration shown in Fig. 11.4(a) provides a direct implementation of the feedback loop of Fig. 11.1.

(a) Assume that the op amp has infinite input resistance and zero output resistance. Find expressions for the feedback factor β and the ideal closed-loop gain A_f.

(b) Find β and R_2/R_1 to obtain an ideal closed-loop gain of 10 V/V.

(c) If the open-loop gain $A = 10^4$ V/V, find for the design in (b) the loop gain, the amount of feedback, and the actual value of A_f. By what percentage does A_f deviate from the ideal value?

(d) To what values must β and R_2/R_1 be changed to obtain a closed-loop gain of exactly 10 V/V?

(e) For the modified design in (d), let $V_s = 1$ V. Find the values of V_o, V_f, and V_i.

(f) If for the design in (d) the open-loop gain decreases by 20%, what is the corresponding decrease in A_f?

Example 11.1 *continued*

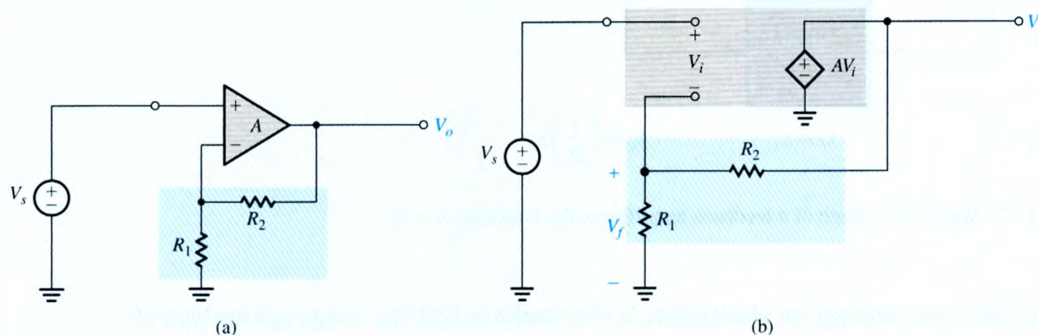


Figure 11.4 **(a)** A noninverting op-amp circuit for Example 11.1. **(b)** The circuit in **(a)** with the op amp replaced with its equivalent circuit.

Solution

(a) To be able to see more clearly the direct correspondence between the circuit in Fig. 11.4(a) and the block diagram in Fig. 11.1, we replace the op amp with its equivalent-circuit model, as shown in Fig. 11.4(b). Since the op amp is assumed to have infinite input resistance and zero output resistance, its model is simply an ideal voltage-controlled voltage source of gain A. From Fig. 11.4(b) we observe that the feedback network, consisting of the voltage divider (R_1, R_2), is connected directly to the output and feeds a signal V_f to the inverting input terminal of the op amp. It is important at this point to note that the zero output resistance of the op amp causes the output voltage to be AV_i irrespective of the values of R_1 and R_2. That is what we meant by the statement that in the block diagram of Fig. 11.1, the feedback network is assumed to not load the basic amplifier. Now we can easily determine the feedback factor β from

$$\beta \equiv \frac{V_f}{V_o} = \frac{R_1}{R_1 + R_2}$$

Let's next examine how V_f is subtracted from V_s at the input side. The subtraction is effectively performed by the differential action of the op amp; by its very nature, a differential-input amplifier takes the difference between the signals at its two input terminals. Observe also that because the input resistance of the op amp is assumed to be infinite, no current flows into the negative input terminal of the op amp and that the feedback network does not load the amplifier at the input side. The ideal closed-loop gain A_f is given by

$$A_f\big|_{\text{ideal}} = \frac{1}{\beta} = 1 + \frac{R_2}{R_1}$$

(b) For $A_f\,|_{\text{ideal}} = 10$ V/V, we require

$$\beta = \frac{1}{A_f\,|_{\text{ideal}}} = 0.1 \text{ V/V}$$

We can obtain the required value of R_2/R_1 from

$$10 = 1 + \frac{R_2}{R_1} \Rightarrow \frac{R_2}{R_1} = 9$$

(c)

$$A\beta = 10^4 \times 0.1 = 1000$$
$$1 + A\beta = 1001 \Rightarrow \simeq 60 \text{ dB}$$
$$A_f = \frac{A}{1 + A\beta} = \frac{10^4}{1001} = 9.990 \text{ V/V}$$

which is 0.1% below the ideal value of 10 V/V, consistent with the loop gain being 1000.

(d)

$$A_f = \frac{A}{1 + A\beta}$$
$$10 = \frac{10^4}{1 + 10^4 \times \beta} \Rightarrow \beta = 0.0999 \text{ V/V}$$
$$\frac{R_1}{R_1 + R_2} = 0.0999 \Rightarrow \frac{R_2}{R_1} = 9.01$$

(e) For $V_s = 1$ V,

$$V_o = A_f V_s = 10 \times 1 = 10 \text{ V}$$
$$V_f = \beta V_o = 0.0999 \times 10 = 0.999 \text{ V}$$
$$V_i = \frac{V_o}{A} = \frac{10}{10^4} = 0.001 \text{ V}$$

(f) If A decreases by 20%, thus becoming

$$A = 0.8 \times 10^4 \text{ V/V}$$

the value of A_f becomes

$$A_f = \frac{0.8 \times 10^4}{1 + 0.8 \times 10^4 \times 0.0999} = 9.9975 \text{ V/V}$$

that is, it decreases by 0.025%, which is less than the percentage change in A by approximately a factor $(1 + A\beta)$.

EXERCISES

11.1 Repeat Example 11.1 (b) to (f) for $A = 100$ V/V.

Ans. (b) 0.1 V/V, 9; (c) 10, 11 or 20.8 dB, 9.091 V/V, −9.1%; (d) 0.09 V/V, 10.11; (e) 10 V, 0.9 V, 0.1 V; (f) 2.44%

11.2 Repeat Example 11.1 (b) to (f) for $A_f|_{\text{ideal}} = 10^3$ V/V. For (d) design for $A_f = 1000$ V/V, and for (e) use $V_s = 0.01$ V.

Ans. (b) 0.001 V/V, 999; (c) 10, 11 or 20.8 dB, 909.1 V/V, −9.1%; (d) 9×10^{-4} V/V, 1110.1; (e) 10 V, 0.009 V, 0.001 V; (f) 2.44%

11.1.5 Summary

Table 11.1, in the online Summary Tables supplement found at www.oup.com/he/sedra-smith8e, presents a summary of the important parameters and formulas that characterize the ideal negative-feedback amplifier structure of Fig. 11.1.

11.2 Some Properties of Negative Feedback

In the introduction we mentioned some of the properties of negative feedback. Here we will consider some of these properties in more detail.

11.2.1 Gain Desensitivity

We demonstrated the effect of negative feedback on desensitizing the closed-loop gain in Example 11.1, where we saw that a 20% reduction in the gain of the basic amplifier gave rise to only a 0.025% reduction in the gain of the closed-loop amplifier. We can derive a formula for sensitivity-reduction as follows:

Assume that β is constant. Taking differentials of both sides of Eq. (11.4) results in

$$dA_f = \frac{dA}{(1+A\beta)^2} \tag{11.15}$$

Dividing Eq. (11.15) by Eq. (11.4) gives

$$\frac{dA_f}{A_f} = \frac{1}{(1+A\beta)}\frac{dA}{A} \tag{11.16}$$

which says that the percentage change in A_f (due to variations in some circuit parameter in the basic amplifier) is smaller than the percentage change in A by a factor equal to the amount of feedback. For this reason, the amount of feedback, $1 + A\beta$, is also known as the **desensitivity factor**.[5]

[5]Since differentiation was used in deriving the expression in Eq. (11.16), it applies for relatively small variations in A. If the variation in A is large, Eq. (11.4) should be utilized to determine the corresponding value of A_f.

EXERCISE

11.3 An amplifier with a nominal gain $A = 1000$ V/V exhibits a gain change of 10% as the operating temperature changes from 25°C to 75°C. If we need to constrain the change to 0.1% by applying negative feedback, what is the largest closed-loop gain possible? If three of these feedback amplifiers are placed in cascade, what are the overall gain and gain variability?
Ans. 10 V/V; 1000 V/V, with a maximum variability of 0.3% over the specified temperature range.

11.2.2 Bandwidth Extension

Consider an amplifier whose high-frequency response is characterized by a single pole. Its gain at mid and high frequencies can be expressed as

$$A(s) = \frac{A_M}{1 + s/\omega_H} \tag{11.17}$$

where A_M denotes the midband gain and ω_H is the upper 3-dB frequency. Application of negative feedback, with a frequency-independent factor β, around this amplifier results in a closed-loop gain $A_f(s)$ given by

$$A_f(s) = \frac{A(s)}{1 + \beta A(s)}$$

Substituting for $A(s)$ from Eq. (11.17) results, after a little manipulation, in

$$A_f(s) = \frac{A_M/(1 + A_M\beta)}{1 + s/\left[\omega_H(1 + A_M\beta)\right]} \tag{11.18}$$

Thus the feedback amplifier will have a midband gain of $A_M/(1 + A_M\beta)$ and an upper 3-dB frequency ω_{Hf} given by

$$\omega_{Hf} = \omega_H(1 + A_M\beta) \tag{11.19}$$

It follows that the upper 3-dB frequency is increased by a factor equal to the amount of feedback. The increase in amplifier bandwidth is accompanied by a decrease in the midband gain by the same factor; namely, the amount of feedback. It follows that the gain–bandwidth product is maintained at a constant value. This important point is illustrated in Fig. 11.5.

The action of negative feedback in extending the amplifier bandwidth should not be surprising: Negative feedback works to minimize the change in gain magnitude, including its change with frequency.

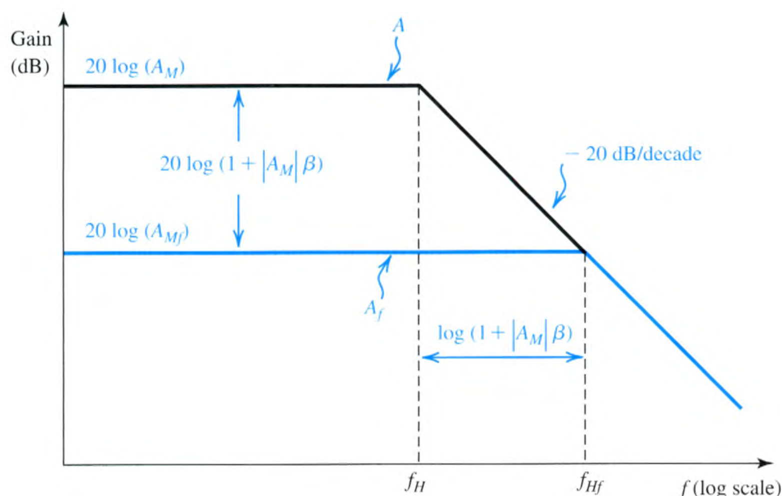


Figure 11.5 Application of negative feedback increases the amplifier bandwidth at the expense of decreasing the midband gain. Both changes are by the factor $(1 + |A_M|\beta)$ which is the amount of feedback. Notice that the gain–bandwidth product remains unchanged.

EXERCISE

11.4 Consider the noninverting op-amp circuit of Example 11.1. Let the open-loop gain A have a low-frequency value of 10^4 V/V and a uniform –20-dB/decade rolloff at high frequencies, with a 3-dB frequency of 100 Hz. Find the low-frequency gain and the upper 3-dB frequency of a closed-loop amplifier with $R_1 = 1$ kΩ and $R_2 = 9$ kΩ.
Ans. 9.99 V/V; 100.1 kHz

11.2.3 Reduction in Nonlinear Distortion

Curve (a) in Fig. 11.6 shows the transfer characteristic v_O versus v_I of an amplifier. As indicated, the characteristic is linear in parts, with the voltage gain changing from 1000 to 100 and then to 0 when the output voltage reaches 1 V and 4 V, respectively. This nonlinear transfer characteristic will result in this amplifier generating a large amount of nonlinear distortion.

The amplifier transfer characteristic can be considerably **linearized** (i.e., made more linear) through the application of negative feedback. The fact that this is possible should not be too surprising, since we have already seen that negative feedback reduces the dependence of the overall closed-loop amplifier gain on the open-loop gain of the basic amplifier. Thus large changes in open-loop gain (1000 to 100 in this case) give rise to much smaller corresponding changes in the closed-loop gain.

To illustrate, let us apply negative feedback with $\beta = 0.01$ to the amplifier whose open-loop voltage transfer characteristic is depicted as curve (a) in Fig. 11.6. The resulting

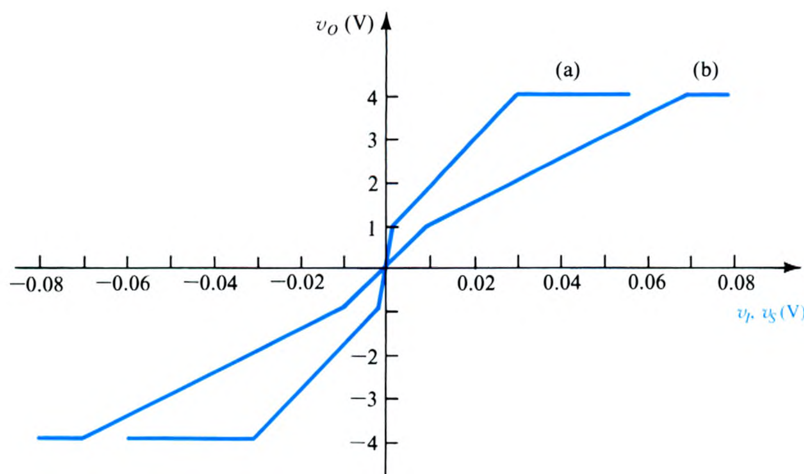


Figure 11.6 Illustrating the application of negative feedback to reduce the nonlinear distortion in amplifiers. Curve **(a)** shows the amplifier transfer characteristic (v_O versus v_I) without feedback. Curve **(b)** shows the characteristic (v_O versus v_S) with negative feedback ($\beta = 0.01$) applied.

transfer characteristic of the closed-loop amplifier, v_O versus v_S, is shown in Fig. 11.6 as curve (b). Here the slope of the steepest segment is

$$A_{f1} = \frac{1000}{1 + 1000 \times 0.01} = 90.9$$

and the slope of the next segment is

$$A_{f2} = \frac{100}{1 + 100 \times 0.01} = 50$$

Thus the factor of 10 change in slope has been considerably reduced. The price paid, of course, is a reduction in voltage gain. Thus if the overall gain has to be restored, a preamplifier should be added. This preamplifier should not present a severe nonlinear-distortion problem, since it will be dealing with smaller signals.

Finally, we note that negative feedback can do nothing at all about amplifier saturation, since in saturation the gain is very small (almost zero) and hence the amount of feedback is almost unity.

Example 11.2

Given the availability of amplifiers with the voltage transfer characteristic shown in Fig. 11.7(a), design an amplifier with a gain of 200 V/V and a linear transfer characteristic for input signals as large as 5 mV.

Solution

The solution consists of cascading one of the given amplifiers (gain of 100) with another to which negative feedback is applied to obtain a closed-loop gain of 2. The feedback factor β for the second amplifier can

Example 11.2 *continued*

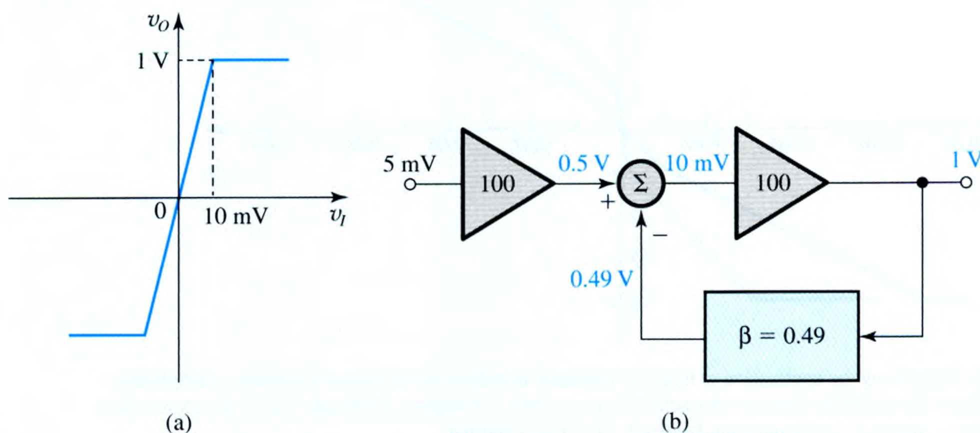


Figure 11.7 **(a)** Transfer characteristic of available amplifier. **(b)** Block diagram of an amplifier with a gain of 200 V/V that is linear for input signals as large as 5 mV.

be found from

$$A_f = \frac{A}{1+A\beta}$$

by substituting $A_f = 2$ and $A = 100$,

$$2 = \frac{100}{1+100\beta}$$

which results in $\beta = 0.49$.

Figure 11.7(b) shows the overall amplifier together with the signal levels at various nodes corresponding to the 5-mV maximum input. Observe that both amplifiers in the cascade are operating in their linear regions.

11.3 The Feedback Voltage Amplifier

Based on the quantity to be amplified (voltage or current) and on the desired form of output (voltage or current), amplifiers can be classified into four categories. We discussed these categories in Chapter 1. In this section we study the most common amplifier type: the voltage amplifier. We begin by identifying the appropriate configuration for applying negative feedback to a voltage amplifier. Then, we present a method for analyzing the feedback voltage amplifier. The method makes use of the loop gain $A\beta$, which we learned how to find in Section 11.1.3.

11.3.1 The Series–Shunt Feedback Topology

Voltage amplifiers are intended to amplify an input voltage signal and provide an output voltage signal. The voltage amplifier is essentially a voltage-controlled voltage source. The input resistance must be high, and the output resistance must be low. Since the signal source is

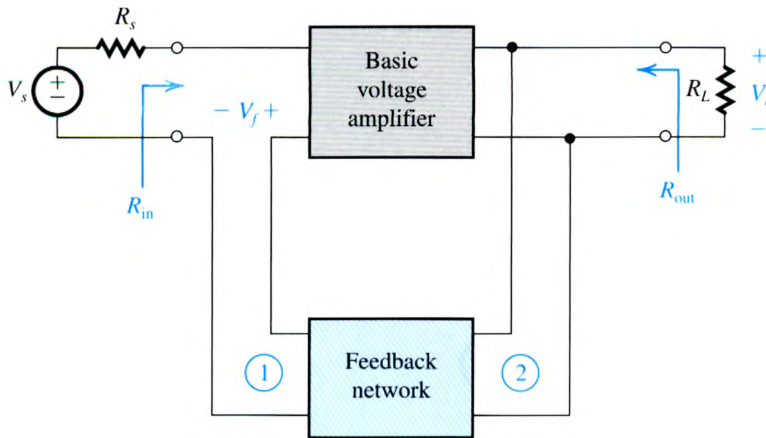


Figure 11.8 Block diagram of a feedback voltage amplifier. Here the appropriate feedback topology is series–shunt.

essentially a voltage source, it is appropriately represented in terms of a Thévenin equivalent circuit. As the output quantity of interest is the output voltage, the feedback network should *sample* the output *voltage*, just as a voltmeter measures a voltage. Also, because of the Thévenin representation of the source, the feedback signal x_f should be a *voltage* that can be *mixed* with the source voltage in *series*.

From the discussion above, it follows that the most suitable feedback topology for the voltage amplifier is the **voltage-mixing, voltage-sampling** one shown in Fig. 11.8. Because of the series connection at the input and the parallel or shunt connection at the output, this feedback topology is also known as **series–shunt feedback**. As we will show, this topology not only stabilizes the voltage gain V_o/V_s but also results in a higher input resistance R_{in} (intuitively, a result of the series connection at the input) and a lower output resistance R_{out} (intuitively, a result of the parallel connection at the output), which are desirable properties for a voltage amplifier.

The increased input resistance results because V_f subtracts from V_s, resulting in a smaller signal V_i at the input of the basic amplifier. The lower V_i, in turn, causes the input current to be smaller, with the result that the resistance seen by V_s will be larger. We will derive a formula for the input resistance of the feedback voltage amplifier in the next section.

The decreased output resistance results because the feedback works to keep V_o as constant as possible. Thus if the current drawn from the amplifier output changes by ΔI_o, the change ΔV_o in V_o will be lower than it would have been without the feedback. Thus the output resistance $\Delta V_o/\Delta I_o$ will be lower than that of the open-loop amplifier. In the following section we will derive an expression for the output resistance of the feedback voltage amplifier.

11.3.2 Examples of Series–Shunt Feedback Amplifiers

Three examples of series–shunt feedback amplifiers are shown in Fig. 11.9. The amplifier in Fig. 11.9(a) is the familiar noninverting op-amp configuration. The feedback network, composed of the voltage divider (R_1, R_2), develops a voltage V_f that is applied to the negative input terminal of the op amp. The subtraction of V_f from V_s is achieved by the differencing action of the op-amp differential input. For the feedback to be negative, V_f must be of the same polarity as V_s, resulting in a smaller signal at the input of the basic amplifier. To confirm

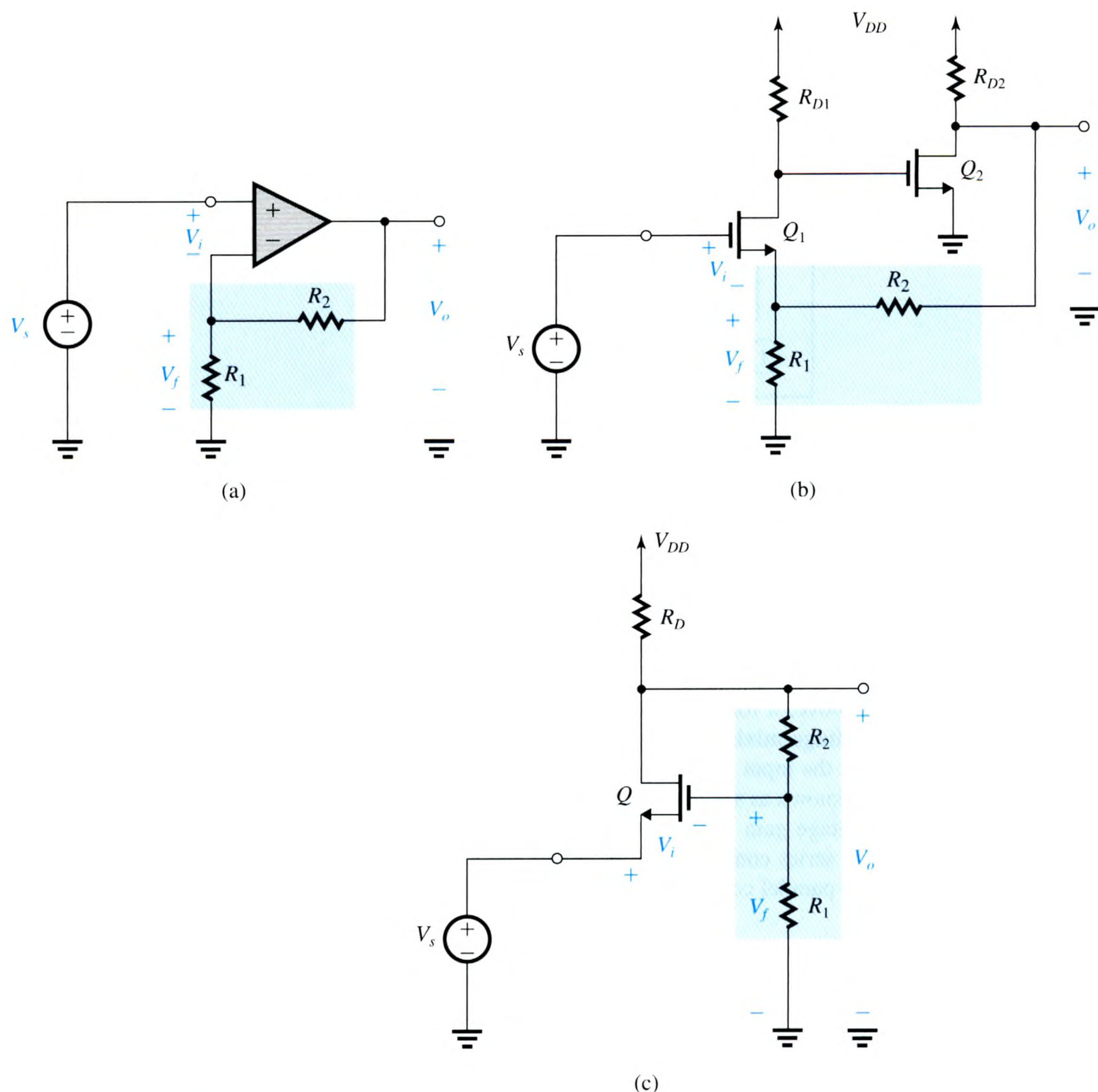


Figure 11.9 Examples of a feedback voltage amplifier. All these circuits use series–shunt feedback. Note that the dc bias circuits are only partially shown.

that this is true, we follow the signal around the loop, as follows: As V_s increases, V_o increases and the voltage divider causes V_f to increase. Thus the change in V_f is of the same polarity as the change in V_s, and the feedback is negative.

The second feedback voltage amplifier, shown in Fig. 11.9(b), utilizes two MOSFET amplifier stages in cascade. The output voltage V_o is sampled by the feedback network composed of the voltage divider (R_1, R_2), and the feedback signal V_f is fed to the source terminal of Q_1. The subtraction is implemented by applying V_s to the gate of Q_1 and V_f to its source, with the result that the signal at this amplifier input $V_i = V_{gs} = V_s - V_f$. To confirm that the feedback is negative, let V_s increase. The drain voltage of Q_1 will decrease, and since

this is applied to the gate of Q_2, its drain voltage V_o will increase. The feedback network will then cause V_f to increase, which is the same polarity initially assumed for the change in V_s. Thus the feedback is indeed negative.

The third example of series–shunt feedback, shown in Fig. 11.9(c), uses a CG transistor Q with a fraction V_f of the output voltage V_o fed back to the gate through a voltage divider (R_1, R_2). Notice that the subtraction of V_f from V_s is effected by applying V_s to the source, so the input V_i to the CG amplifier is $V_s - V_f$. As usual, however, we must check the polarity of the feedback: If V_s increases, V_d (which is V_o) will increase and V_f will likewise increase. Thus V_f and V_s change in the same direction, verifying that the feedback is negative.

FEEDBACK: A HISTORICAL NOTE

The idea of feedback as an element of self-regulating behavior dates back to the eighteenth century, but the term itself did not appear until the 1860s, in the context of a discussion on economics. Still later, in 1909, Karl Ferdinand Braun, a German physicist working at the University of Strasbourg, referred publicly to feedback as an undesired coupling between components of a vacuum-tube electronic system. The occasion was the lecture Braun delivered as a recipient of the Nobel Prize in Physics, shared with Guglielmo Marconi (often solely credited as the inventor of radio).

In 1927, Harold Black, at Bell Labs, invented the negative-feedback amplifier, which he detailed in a seminal paper, "Stabilized Feedback Amplifiers," published in 1934. This invention was motivated by the need to provide low-distortion amplifiers that could be concatenated in long-distance transcontinental telephone circuits.

11.3.3 Analysis of the Feedback Voltage Amplifier

Given a feedback voltage amplifier circuit like any of the three shown in Fig. 11.9, or more generally like the one shown in block diagram form in Fig. 11.8, our goal is to analyze the circuit to determine the closed-loop gain $A_f \equiv V_o/V_s$, the input resistance R_{in}, and the output resistance R_{out}. Our natural inclination would be to use the formulas derived in Section 11.1. However, this is not possible directly because the analysis of Section 11.1 assumes that the feedback network does not load the basic amplifier, and unfortunately, this assumption does not hold in most practical amplifier circuits. As shown in Fig. 11.9, the feedback network is a simple resistive circuit that obviously loads the basic amplifier. As an example, in the circuit of Fig. 11.9(b), the values of the resistances R_2 and R_1, which comprise the feedback network, affect the gain of the common-source stage Q_2, which is part of the basic amplifier. Also, the value of the feedback-network resistance R_1 affects the gain of the Q_1 amplifier stage, which is part of the basic amplifier. It follows that we cannot easily disassemble a practical amplifier circuit to determine A and β and thus be able to use the feedback formulas of Sections 11.1 and 11.2. As an alternative, we present the following three-step analysis method:

Step 1: Determine β and $A_f|_{ideal}$ We use the limit case of infinite open-loop gain A (Section 11.1.4). Figure 11.10(a) shows the general feedback voltage amplifier of Fig. 10.8 with the gain of the basic amplifier assumed infinite. This results in a zero voltage across its input terminals, which in turn results in a zero input current. The zero current in the input series loop results in $V_f = V_s$. It is very important to note that port 1 of the feedback network will be operating as an open circuit, and thus β can be determined as shown in Fig. 11.10(b).

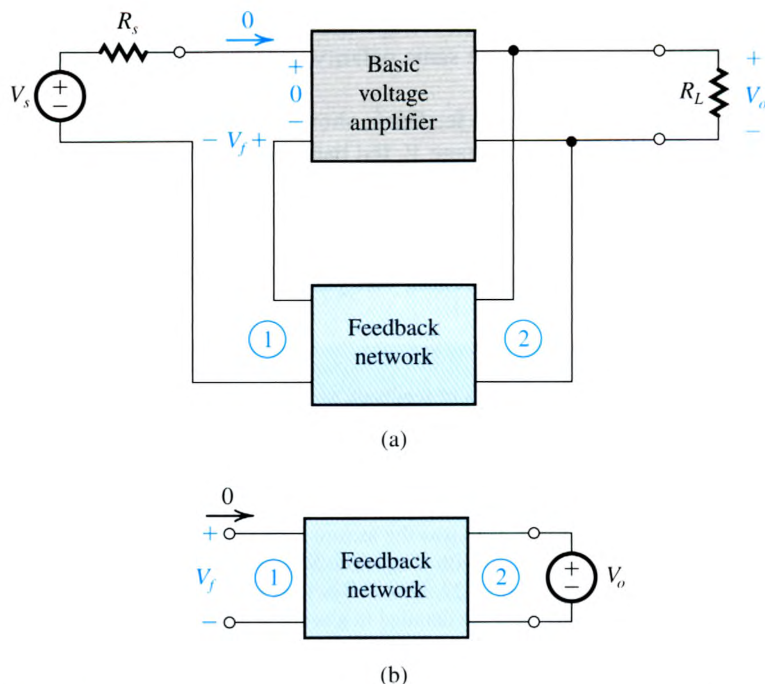


Figure 11.10 **(a)** The general feedback voltage amplifier under the assumption that the open-loop gain A of the basic amplifier is infinite. The resulting 0 V between the input terminals and the correspondingly zero current in the input series loop enable us to determine β as shown in **(b)**.

The ideal value of the closed-loop gain can then be determined as

$$A_f \mid_{\text{ideal}} = \frac{1}{\beta}$$

Step 2: Determine the Loop Gain $A\beta$ The loop gain $A\beta$ can be determined using the method presented in Section 11.1.3. To begin, we break the feedback loop while taking care to not change the conditions that existed in the loop before breaking it. This becomes clear if you refer to Fig. 11.11. The conceptual loop in Fig. 11.11(a) is broken at XX'. As shown in Fig. 11.11(b), we apply a test voltage V_t to the terminals thus created to the left of XX' and, to keep the loop conditions unchanged, we connect to the terminals to the right of XX' an impedance Z_t whose value is equal to the impedance previously seen looking to the left of XX'. The loop gain is then determined from

$$A\beta = -\frac{V_r}{V_t}$$

Whenever possible, we should break the loop at a location where Z_t is infinite.

Step 3: Determine the Closed-Loop Gain A_f Using the value of β determined in step 1 and the value of $A\beta$ determined in step 2, we can find the value of the open-loop gain A.

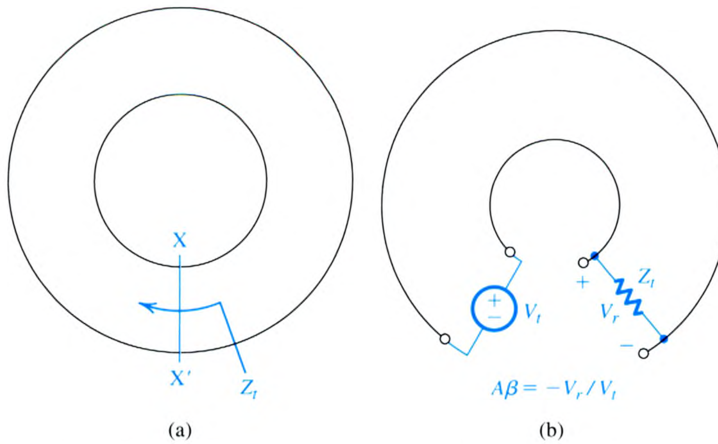


Figure 11.11 Breaking the conceptual feedback loop in (**a**) to determine the loop gain requires the termination of the loop as shown in (**b**), to ensure that the loop conditions do not change.

Finally, we can find the value of the closed-loop gain A_f from

$$A_f = \frac{A}{1 + A\beta}$$

We can then compare this value to the ideal (upper bound) value we found in step 1. The difference should be approximately $(-100/A\beta)\%$.

We will use the next two examples to show how this analysis method is applied.

Example 11.3

For the series–shunt feedback amplifier of Fig. 11.9(b), neglect the MOSFETs' r_o and

(a) find expressions for β and the ideal value of the closed-loop gain A_f.
(b) find the ratio R_2/R_1 that results in an ideal closed-loop gain of 10 V/V. If $R_1 = 1\text{ k}\Omega$, what value must R_2 have?
(c) find an expression for the loop gain $A\beta$.
(d) if $g_{m1} = g_{m2} = 4$ mA/V and $R_{D1} = R_{D2} = 10\text{ k}\Omega$, determine the values of $A\beta$, A, and A_f.

Solution

(a) To determine β and $A_f\,|_{\text{ideal}}$, we assume infinite open-loop gain as shown in Fig. 11.12(a). Notice that the input voltage to the basic amplifier, V_{gs1}, is zero, and correspondingly the current in the source of Q_1 is zero. Thus, the feedback factor β can be found as the ratio of the voltage divider (R_1, R_2),

$$\beta = \frac{V_f}{V_o} = \frac{R_1}{R_1 + R_2}$$

The ideal value of the closed-loop gain $A_f = V_o/V_s$ is given by

$$|A_f|_{\text{ideal}} = \frac{1}{\beta} = 1 + \frac{R_2}{R_1}$$

Example 11.3 *continued*

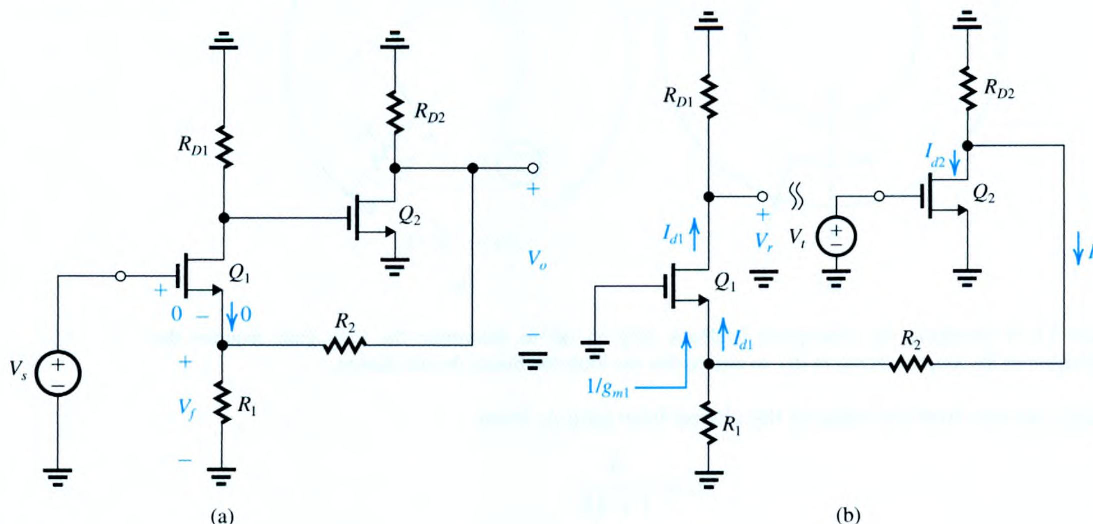


Figure 11.12 **(a)** The feedback voltage amplifier with the open-loop gain A assumed infinite to determine β and $A_f|_{\text{ideal}}$. **(b)** Determining the loop gain $A\beta$.

(b) For A_f to have an ideal value of 10,

$$10 = 1 + \frac{R_2}{R_1}$$

Thus,

$$\frac{R_2}{R_1} = 9$$

For $R_1 = 1\ \text{k}\Omega$, $R_2 = 9\ \text{k}\Omega$.

(c) To determine the loop gain, we set $V_s = 0$. Examining the feedback loop reveals that it is most convenient to break the loop at the connection between the drain of Q_1 and the gate of Q_2. This is because of the infinite input impedance at the gate of Q_2. The resulting circuit is shown in Fig. 11.12(b), for which the analysis to determine $A\beta \equiv -V_r/V_t$ proceeds as follows:

$$I_{d2} = g_{m2} V_t$$

$$I_1 = -I_{d2} \frac{R_{D2}}{R_{D2} + R_2 + \left(R_1 \left\| \frac{1}{g_{m1}} \right.\right)}$$

$$I_{d1} = I_1 \frac{R_1}{R_1 + \frac{1}{g_{m1}}}$$

$$V_r = I_{d1} R_{D1}$$

Combining these four equations results in

$$A\beta \equiv -\frac{V_r}{V_t} = (g_{m1}R_{D1})(g_{m2}R_{D2})\frac{1}{1+g_{m1}R_1}\frac{R_1}{R_{D2}+R_2+\left(R_1 \left\| \frac{1}{g_{m1}}\right.\right)}$$

(d)

$$A\beta = 4 \times 10 \times 4 \times 10 \times \frac{1}{1+4\times 1} \times \frac{1}{10+9+(1\|\frac{1}{4})}$$
$$= 16.67$$

Since

$$\beta = \frac{R_1}{R_1+R_2} = \frac{1}{1+9} = 0.1$$
$$A = \frac{A\beta}{\beta} = \frac{16.67}{0.1} = 166.7 \text{ V/V}$$

and

$$A_f = \frac{A}{1+A\beta} = \frac{166.7}{1+16.67} = \frac{166.7}{17.67} = 9.43 \text{ V/V}$$

which, given that the loop gain is only moderately high (16.67), is reasonably close to the ideal value of 10 V/V. In fact, A_f is only 5.7% below the ideal value of 10 V/V, which compares well to the expected deviation of $100/16.67 \simeq 6\%$.

Example 11.4

In the series–shunt feedback amplifier of Fig. 11.13(a), the op amp has an input resistance R_{id}, an open-circuit voltage gain μ, and an output resistance r_o. Find expressions for β, the ideal value of $A_f \equiv V_o/V_s$, and the loop gain $A\beta$. For $\mu = 10^4$ V/V, $R_{id} = 100$ kΩ, $r_o = 1$ kΩ, $R_L = 2$ kΩ, $R_1 = 1$ kΩ, $R_2 = 100$ kΩ, and $R_s = 10$ kΩ, find β, $A_f|_{\text{ideal}}$, $A\beta$, A, and A_f.

Solution

To determine β we assume the open-loop gain is infinite, which results in a zero voltage between the op amp's input terminals and thus the current into the input terminals will also be zero, as shown in Fig. 11.13(b). The zero current in the input series loop enables us to determine the feedback factor β as the ratio of the voltage divider (R_1, R_2),

$$\beta = \frac{R_1}{R_1+R_2}$$

and the ideal value of A_f is

$$A_f|_{\text{ideal}} = \frac{1}{\beta} = 1 + \frac{R_2}{R_1}$$

For the numerical values given, $\beta = 0.0099$ and $A_f|_{\text{ideal}} = 101$ V/V.

Example 11.4 *continued*

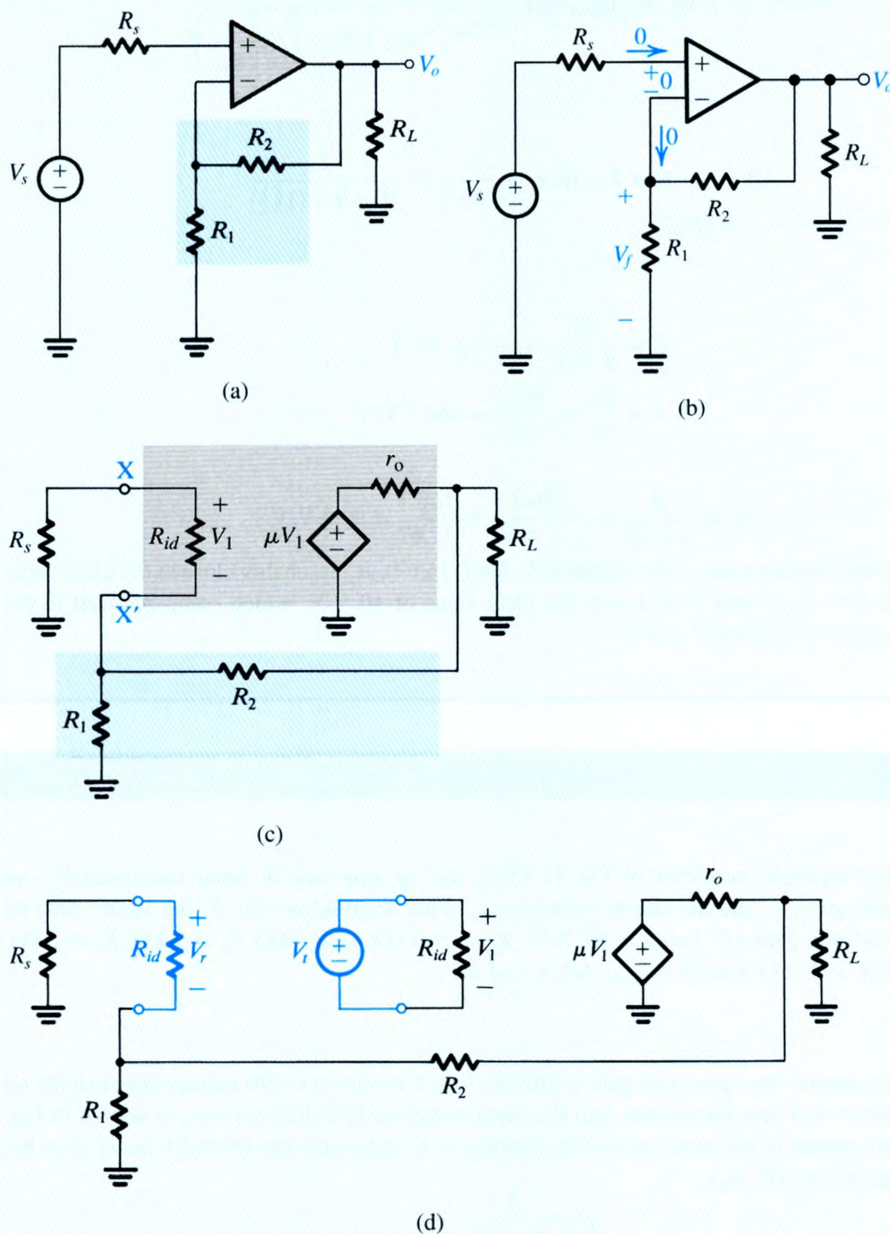


Figure 11.13 Example 11.4. (**a**) A series–shunt feedback amplifier; (**b**) the feedback amplifier with the open-loop gain A assumed infinite; (**c**) the feedback loop obtained by setting $V_s = 0$ and replacing the op amp with its equivalent-circuit model; (**d**) breaking the feedback loop to determine the loop gain $A\beta = -V_r/V_t$.

To determine the loop gain, we set $V_s = 0$ and replace the op amp with its equivalent-circuit model. The resulting circuit is shown in Fig. 11.13(c). Next, we break the loop to apply a test voltage V_t while terminating the loop at the break in an impedance equal to that seen prior to breaking the loop. The resulting circuit is shown in Fig. 11.13(d), where the loop has been broken at the input terminals of the op amp and the left-hand-side terminals are connected to a resistance equal to R_{id}. Analysis of the circuit to determine $A\beta \equiv -V_r/V_t$ involves repeated application of the voltage divider rule, resulting in

$$A\beta = \mu \frac{\{R_L \| [R_2 + R_1 \| (R_{id} + R_s)]\}}{\{R_L \| [R_2 + R_1 \| (R_{id} + R_s)]\} + r_o} \times$$

$$\frac{[R_1 \| (R_{id} + R_s)]}{[R_1 \| (R_{id} + R_s)] + R_2} \times \frac{R_{id}}{R_{id} + R_s}$$

For the numerical values given, $A\beta = 60$, $A = A\beta/\beta = 5960$ V/V, and $A_f = A/(1 + A\beta) = 99.3$ V/V.

Video Example VE 11.1 Design and Analysis of a Feedback Voltage Amplifier Using Loop Gain

Figure VE11.1 shows a series–shunt feedback amplifier known as a "feedback triple." All three MOSFETs are biased to operate at $g_m = 5$ mA/V. You may neglect their r_o's.

(a) Select a value for R_F that results in a closed-loop gain that is ideally 10 V/V.

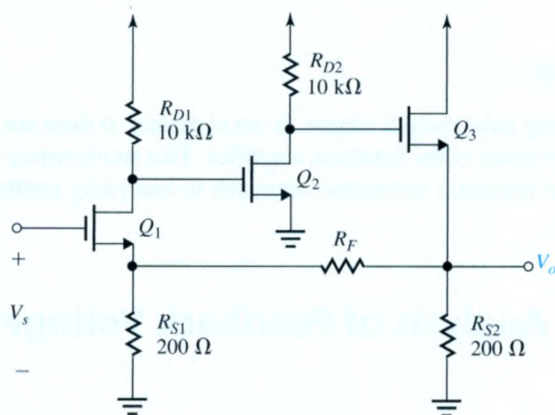


Figure VE11.1 Circuit for Video Example 11.1.

(b) Determine the loop gain $A\beta$ and hence the value of A_f. By what percentage does A_f differ from the ideal value you designed for? How can you adjust the circuit to make A_f equal to 10?



Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

VE 11.1

Related end-of-chapter problem: 11.28

EXERCISE

11.5 For the feedback voltage amplifier of Fig. 11.9(c):

(a) Find an expression for β.

(b) Neglecting the MOSFET r_o, find an expression for the loop gain $A\beta$. (*Hint*: Break the loop at the gate of Q.)

(c) Find an expression for the open-loop gain A.

(d) For $g_m = 4$ mA/V, $R_D = 10$ kΩ, $R_1 = 20$ kΩ, and $R_2 = 80$ kΩ, find the values of β, $A\beta$, A, and the closed-loop gain A_f. What would A_f be if $A\beta$ were much greater than unity?

Ans. (a) $\dfrac{R_1}{R_1+R_2}$; (b) $g_m\dfrac{R_DR_1}{R_D+R_1+R_2}$; (c) $g_m\dfrac{R_D(R_1+R_2)}{R_D+R_1+R_2}$; (d) 0.2, 7.27, 36.36 V/V, 4.4 V/V, 5 V/V.

11.3.4 A Final Remark

The analysis method using loop gain, though simple, is not complete: It does not allow us to find the input and output resistances of the feedback amplifier. This shortcoming is remedied in the next section, where we present a systematic approach to analyzing feedback voltage amplifiers.

11.4 Systematic Analysis of Feedback Voltage Amplifiers

In this section we provide a systematic procedure for analyzing feedback voltage amplifiers. The procedure essentially disassembles a given feedback voltage-amplifier circuit to obtain the "A circuit," from which we can determine the open-loop gain A and other parameters of the open-loop amplifier, such as the input and output resistances, and the "β circuit" from which we can find the value of the feedback factor β. We can then use the feedback formulas to determine the characteristic parameters of the feedback amplifier, such as the closed-loop gain A_f and the input and output resistances with feedback.

Our approach will be to first consider the *ideal* case in which the feedback network does *not* load the basic amplifier. Then, we consider the practical case in which not only does the feedback network load the basic amplifier, but also there is a finite source resistance R_s and a finite load resistance R_L.

11.4.1 The Ideal Case

As mentioned before, series–shunt is the appropriate feedback topology for a voltage amplifier. The ideal structure of the series–shunt feedback amplifier, shown in Fig. 11.14(a), consists of a *unilateral* open-loop amplifier (the A circuit) and an ideal voltage-sampling, voltage-mixing feedback network (the β circuit). The A circuit has an input resistance R_i, an open-circuit voltage gain A, and an output resistance R_o. We assume that the source is ideal with a zero resistance and that there is no load resistance. Furthermore, note that the β circuit does *not* load the A circuit; that is, connecting the β circuit does not change the value of A (defined as $A \equiv V_o/V_i$).

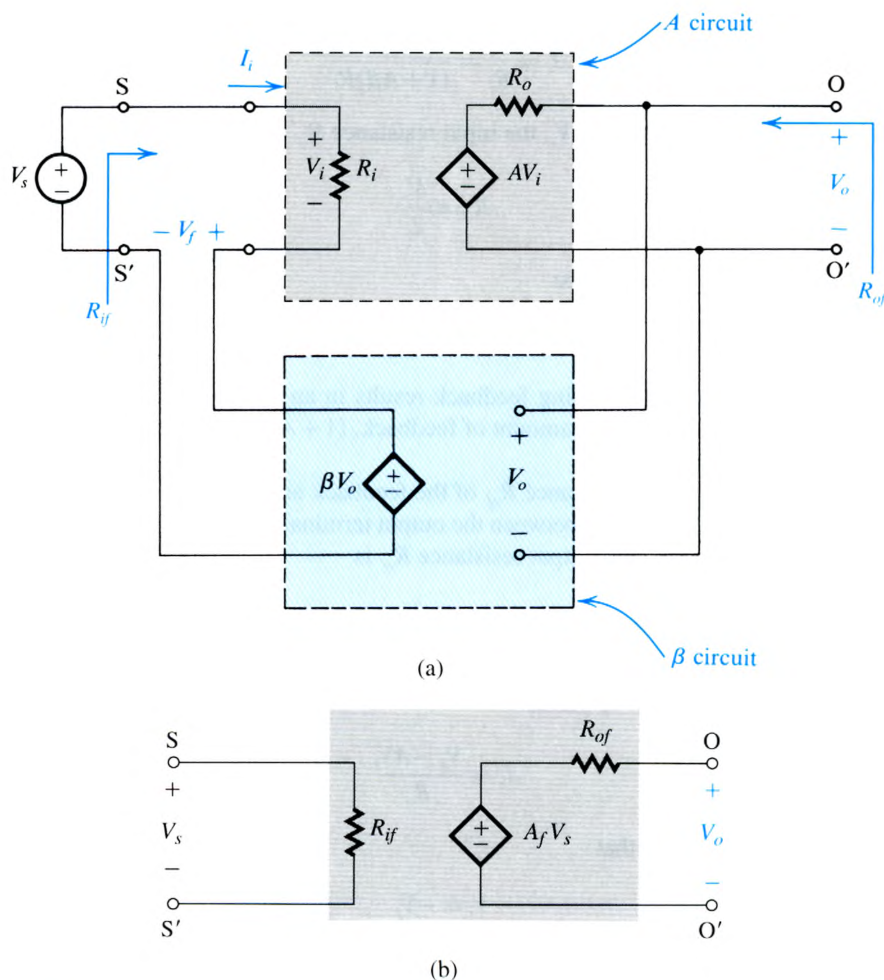


Figure 11.15 The series–shunt feedback amplifier: **(a)** ideal structure; **(b)** equivalent circuit.

The circuit of Fig. 11.14(a) exactly follows the ideal feedback model of Fig. 11.1. Therefore, the closed-loop voltage gain A_f is given by

$$A_f \equiv \frac{V_o}{V_s} = \frac{A}{1+A\beta} \tag{11.20}$$

The equivalent-circuit model of the series–shunt feedback amplifier is shown in Fig. 11.14(b). Notice that A_f is the open-circuit voltage gain of the feedback amplifier, R_{if} is its input resistance, and R_{of} is its output resistance. We can derive expressions for R_{if} and R_{of} as follows.

For R_{if}, refer to the input loop of the circuit in Fig. 11.14(a). The series mixing subtracts V_f from V_s and thus reduces V_i by a factor equal to the amount of feedback (Eq. 11.10),

$$V_i = \frac{V_s}{1+A\beta}$$

Thus the input current I_i becomes

$$I_i = \frac{V_i}{R_i} = \frac{V_s}{(1+A\beta)R_i} \tag{11.21}$$

Since I_i is the current drawn from V_s, the input resistance R_{if} can be expressed as

$$R_{if} \equiv \frac{V_s}{I_i}$$

and using Eq. (11.21) is found to be

$$R_{if} = (1+A\beta)R_i \tag{11.22}$$

Thus, as expected, the series-mixing feedback results in an increase in the amplifier input resistance by a factor equal to the amount of feedback, $(1+A\beta)$, a highly desirable property for a voltage amplifier.

To determine the output resistance R_{of} of the feedback amplifier in Fig. 11.14(a), we set $V_s = 0$ and apply a test voltage V_x between the output terminals, as shown in Fig. 11.15. If the current drawn from V_x is I_x, the output resistance R_{of} is

$$R_{of} \equiv \frac{V_x}{I_x} \tag{11.23}$$

An equation for the output loop yields

$$I_x = \frac{V_x - AV_i}{R_o} \tag{11.24}$$

From the input loop we see that

$$V_i = -V_f$$

Now $V_f = \beta V_o = \beta V_x$; thus,

$$V_i = -\beta V_x$$

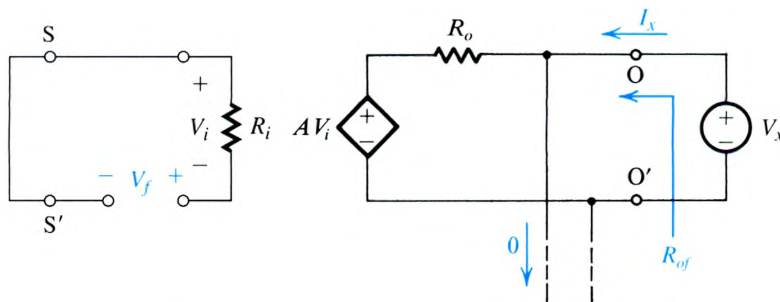


Figure 11.16 Determining the output resistance of the feedback amplifier of Fig. 11.14(a): $R_{of} = V_x/I_x$.

which when substituted in Eq. (11.24) gives

$$I_x = \frac{V_x(1+A\beta)}{R_o}$$

Substituting this value of I_x into Eq. (11.23) provides the following expression for R_{of},

$$R_{of} = \frac{R_o}{1+A\beta} \tag{11.25}$$

As expected, the shunt sampling (or voltage sampling) at the output results in a decrease in the amplifier output resistance by a factor equal to the amount of negative feedback, $(1+A\beta)$, another highly desirable property for a voltage amplifier.

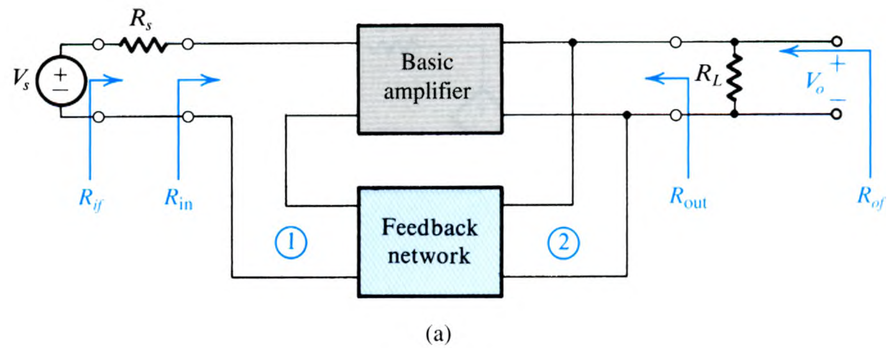
Although perhaps not entirely obvious, the reduction of the output resistance is a result only of the method of sampling the output and does not depend on the method of mixing.

11.4.2 The Practical Case

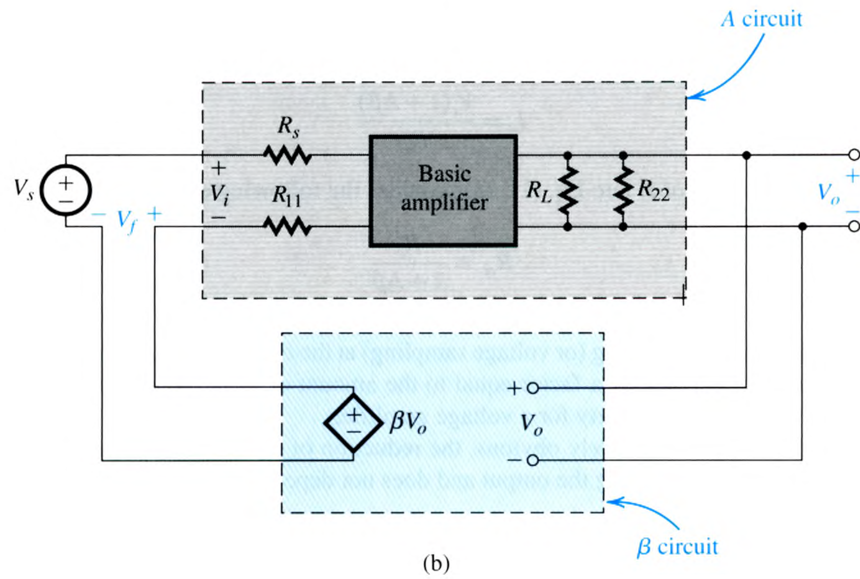
In a practical series–shunt feedback amplifier, the feedback network will not be an ideal voltage-controlled voltage source. Instead, as we can see in Fig. 11.9, the feedback network is usually resistive and hence will load the basic amplifier and thus affect the values of A, R_i, and R_o. There will also be finite source and load resistances that will in turn affect these three parameters. Thus the problem we have is this: Given a series–shunt feedback amplifier represented by the block diagram of Fig. 11.16(a), find the A circuit and the β circuit.

The problem in essence is to represent the general feedback voltage amplifier of Fig. 11.16(a) with the ideal structure of Fig. 11.14(a). The solution is presented, without derivation, in Fig. 11.16(b). We make the following observations.

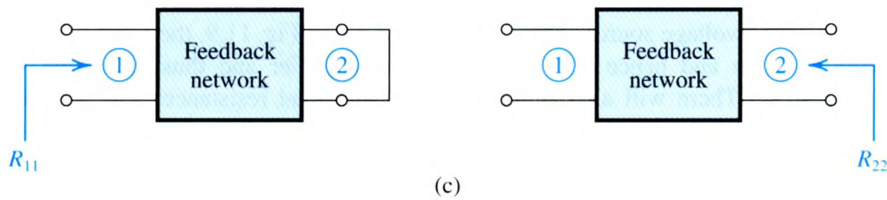
1. The A circuit is obtained by augmenting the basic amplifier at the input with the source resistance R_s and a resistance R_{11}, and at the output with the load resistance R_L and a resistance R_{22}. Resistances R_{11} and R_{22} represent the loading effect of the feedback network on the basic amplifier at the input and the output, respectively.
2. Resistance R_{11} is the resistance looking into port 1 of the feedback network while port 2 is short-circuited. Resistance R_{22} is the resistance looking into port 2 of the



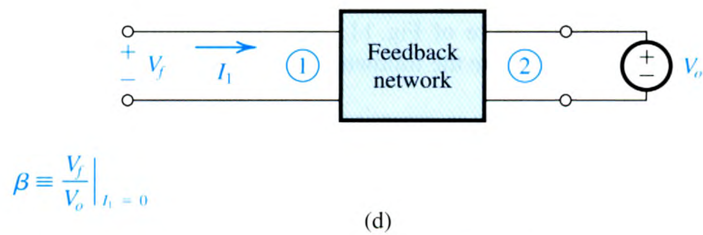
(a)



(b)



(c)



(d)

Figure 11.17 **(a)** Block diagram of a practical series–shunt feedback amplifier. **(b)** The circuit in **(a)** represented by the ideal structure of Fig. 11.12(a). **(c)** Definition of R_{11} and R_{22}. **(d)** Determination of the feedback factor β. **(e)** The A circuit, showing the open-loop resistances R_i and R_o.

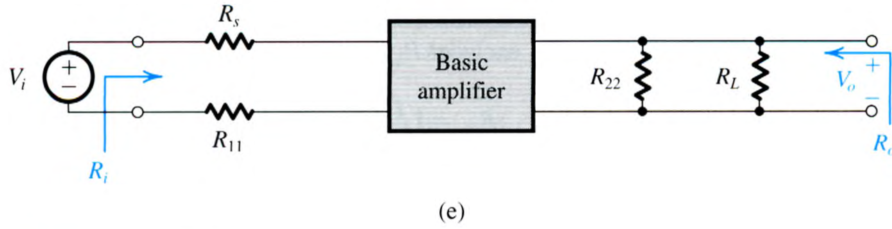


Figure 11.16 *continued*

feedback network while port 1 is open-circuited. These definitions are illustrated in Fig. 11.16(c). Since the feedback network is connected in shunt with the output, shorting its port 2 destroys the feedback. Similarly, because the feedback network is connected in series with the input, opening its port 1 destroys the feedback. It follows that we can find the loading effect of the feedback network by looking into its appropriate port while the other port is open-circuited or short-circuited so as to destroy the feedback. A simple rule to remember: If the connection is *sh*unt, *sh*ort it; if *se*ries, *se*ver it.

3. The feedback factor β is the transmission from port 2 to port 1 of the feedback network, with port 1 open-circuited (which destroys the feedback). Fig. 11.16(c) shows that

$$\beta \equiv \left.\frac{V_f}{V_o}\right|_{I_1=0} \tag{11.26}$$

Note that this expression for β is the same as the one we found using the method outlined in Section 11.3.3.

4. The open-loop gain A can be found from the A circuit in Fig. 11.16(e) as $A = \dfrac{V_o}{V_i}$

5. The values of A and β can be used to find the closed-loop gain A_f,

$$A_f \equiv \frac{V_o}{V_s} = \frac{A}{1+A\beta}$$

6. The open-loop input resistance R_i and output resistance R_o can be found from the A circuit [see Fig. 11.16(e)]. These values can be used to determine the input and output resistances with feedback,

$$R_{if} = R_i(1+A\beta)$$
$$R_{of} = R_o/(1+A\beta)$$

From Fig. 11.16(a) we see that R_{if} is the resistance seen by the ideal signal source V_s. The *actual* input resistance of the feedback amplifier R_{in} excludes R_s [See Fig. 11.16(a)] and is found from R_{if},

$$R_{in} = R_{if} - R_s \tag{11.27}$$

Similarly, R_{of} is the output resistance of the feedback amplifier including R_L. The *actual* output resistance R_{out} excludes R_L [see Fig. 11.16(a)] and is found from R_{of},

$$R_{out} = 1 \Big/ \left(\frac{1}{R_{of}} - \frac{1}{R_L} \right) \tag{11.28}$$

A final and important note: The representation in Fig. 11.16(b) is only *approximately* equivalent to the original circuit in Fig. 11.16(a). The approximation is a result of neglecting the small forward transmission in the feedback network relative to the much larger forward transmission in the basic amplifier. Also, recall that we continue to assume that the basic amplifier is unilateral—that is, it does not have internal feedback; all the feedback occurs in the feedback network and is represented by the feedback factor β.

Example 11.5 Feedback Voltage Amplifier Using an Op Amp

Figure 11.17(a) shows an op amp connected in the noninverting configuration. The op amp has an open-circuit voltage gain μ, a differential input resistance R_{id}, and an output resistance r_o. Recall that in our analysis of op-amp circuits in Chapter 2, we neglected the effects of R_{id} (assumed it to be infinite) and of r_o (assumed it to be zero). Here we wish to use the feedback method to analyze the circuit taking both R_{id} and r_o into account. Find A, β, the ideal and actual value of the closed-loop gain V_o/V_s, the input resistance R_{in}, and the output resistance R_{out}. Let $\mu = 10^4$, $R_{id} = 100\ \text{k}\Omega$, $r_o = 1\ \text{k}\Omega$, $R_L = 2\ \text{k}\Omega$, $R_1 = 1\ \text{k}\Omega$, $R_2 = 100\ \text{k}\Omega$, and $R_s = 10\ \text{k}\Omega$. Note that this circuit was analyzed in Example 11.4 using the loop-gain method; where appropriate, compare results.

Solution

We see that the feedback network consists of R_2 and R_1. This network samples the output voltage V_o and provides a voltage signal (across R_1) that is mixed in series with the input source V_s.

The A circuit can be easily found using the rules of Fig. 11.16, and is shown in Fig. 11.17(b). Notice that the loading effect of the feedback network at the input side is obtained by short-circuiting port 2 of the feedback network (because it is connected in shunt) and looking into port 1, with the result that $R_{11} = R_1 \| R_2$. The loading effect of the feedback network at the output side is found by open-circuiting port 1 (because it is connected in series) and looking into port 2, with the result that $R_{22} = R_2 + R_1$. For the resulting A circuit in Fig. 11.17(b), we can write by inspection:

$$A \equiv \frac{V_o}{V_i} = \mu \frac{R_L \| (R_1 + R_2)}{[R_L \| (R_1 + R_2)] + r_o} \frac{R_{id}}{R_{id} + R_s + (R_1 \| R_2)}$$

For the values given, we find that A = 5967 V/V, which is very close to the value of 5960 found in Example 11.4.

The circuit for determining β is shown in Fig. 11.17(c), from which we obtain

$$\beta \equiv \frac{V_f}{V_o} = \frac{R_1}{R_1 + R_2} = 0.0099\ \text{V/V}$$

The ideal value of A_f is

$$A_f \Big|_{\text{ideal}} = \frac{1}{\beta} = 101\ \text{V/V}$$

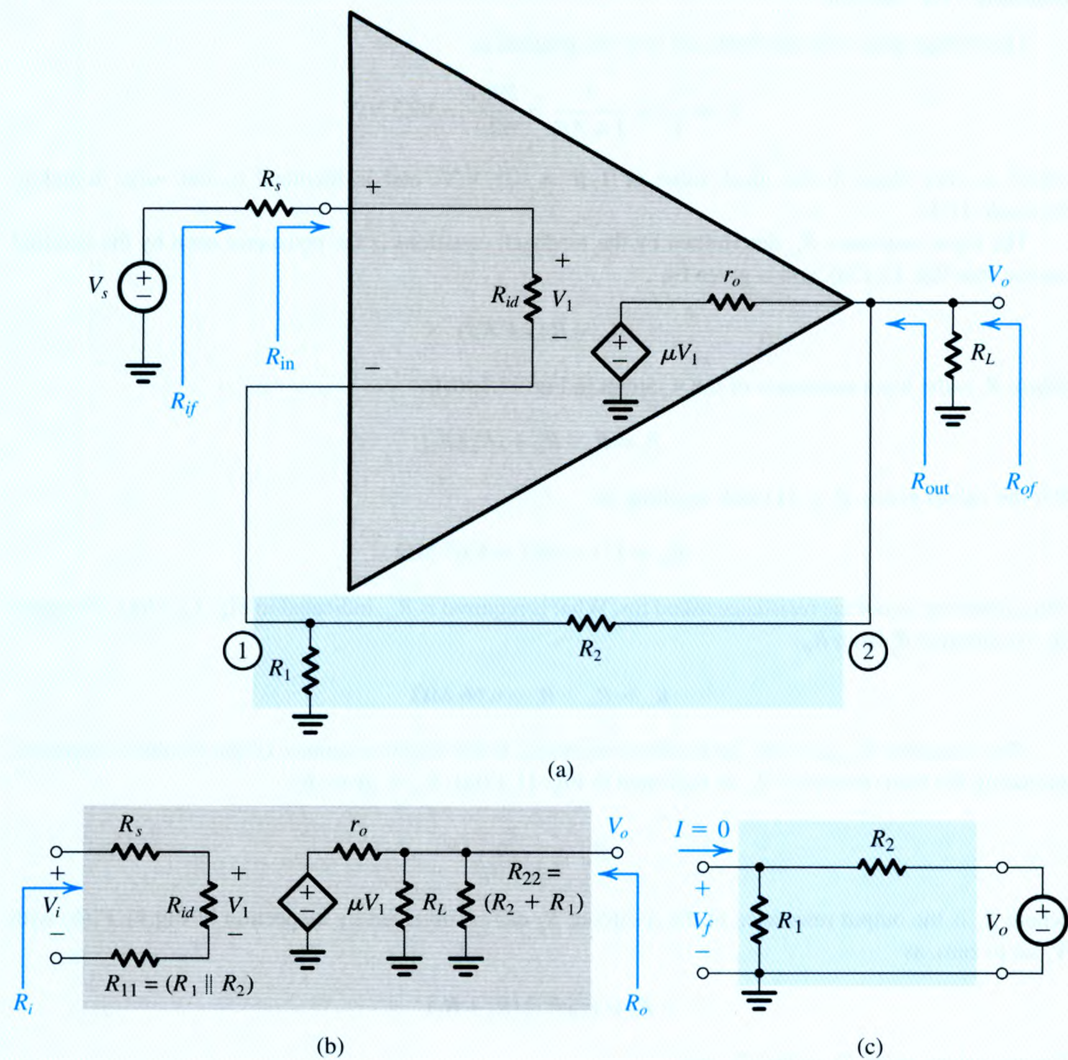


Figure 11.17 Circuits for Example 11.5.

The values of β and $A_f\big|_{\text{ideal}}$ are identical to those found in Example 11.4. The loop gain can now be found as

$$A\beta = 59.1$$

which is very close to the value (60) found in Example 11.4, and the amount of feedback as

$$1 + A\beta = 60.1$$

Example 11.5 *continued*

The voltage gain with feedback can now be obtained as

$$A_f \equiv \frac{V_o}{V_s} = \frac{A}{1+A\beta} = \frac{5967}{60.1} = 99.3 \text{ V/V}$$

which is very close to the ideal value of $1/\beta = 101$ V/V, and is identical to the value found in Example 11.4.

The input resistance R_{if} determined by the feedback equations is the resistance seen by the external source (see Fig. 11.17a), and is given by

$$R_{if} = R_i(1+A\beta)$$

where R_i is the input resistance of the A circuit in Fig. 11.17(b):

$$R_i = R_s + R_{id} + (R_1 \parallel R_2)$$

For the values given, $R_i \simeq 111$ kΩ, resulting in

$$R_{if} = 111 \times 60.1 = 6.67 \text{ M}\Omega$$

This, however, is not the resistance asked for. What is required is R_{in}, indicated in Fig. 11.17(a). To obtain R_{in} we subtract R_s from R_{if}:

$$R_{\text{in}} = R_{if} - R_s = 6.66 \text{ M}\Omega$$

The resistance R_{of} given by the feedback equations is the output resistance of the feedback amplifier, including the load resistance R_L, as indicated in Fig. 11.17(a). R_{of} is given by

$$R_{of} = \frac{R_o}{1+A\beta}$$

where R_o is the output resistance of the A circuit. R_o can be obtained by inspection of Fig.11.17(b) with V_i set to zero, as

$$R_o = r_o \parallel R_L \parallel (R_2 + R_1)$$

For the values given, $R_o \simeq 662\ \Omega$, and

$$R_{of} = \frac{662}{60.1} = 11\ \Omega$$

The resistance asked for, R_{out}, is the output resistance of the feedback amplifier excluding R_L. From Fig. 11.17(a) we see that

$$R_{of} = R_{\text{out}} \parallel R_L$$

Thus

$$R_{\text{out}} \simeq 11\ \Omega$$

Example 11.6 Feedback Voltage Amplifier Using a Pair of MOSFETs

As another example of a series–shunt feedback amplifier, consider the circuit shown in Fig. 11.9(b), which we analyzed in Example 11.3 by determining the loop gain $A\beta$. In this example we want to first analyze the circuit using our systematic procedure and then compare the results to those we got in Example 11.3. For convenience, the circuit is repeated in Fig. 11.18(a). We need to obtain the voltage gain V_o/V_s, input resistance R_{in}, and output resistance R_{out}. Find numerical values for the case $g_{m1} = g_{m2} = 4$ mA/V, $R_{D1} = R_{D2} = 10$ kΩ, $R_1 = 1$ kΩ, and $R_2 = 9$ kΩ. For simplicity, neglect r_o of each of Q_1 and Q_2.

Solution

We identify the feedback network as the voltage divider (R_1, R_2). We find its loading effect at the input by short-circuiting its port 2 (because it is connected in shunt with the output). Then, looking into its port 1, we see $R_1 \parallel R_2$. The loading effect at the output is found by open-circuiting port 1 of the feedback network (because it is connected in series with the input). Then, looking into port 2, we see R_2 in series with R_1. The A circuit will therefore be as shown in Fig. 11.18(b). The gain A is determined as the product of the gain of Q_1 and the gain of Q_2 as follows:

$$A_1 = \frac{V_{d1}}{V_i} = -\frac{R_{D1}}{1/g_{m1} + (R_1 \parallel R_2)} = -\frac{g_{m1}R_{D1}}{1 + g_{m1}(R_1 \parallel R_2)}$$

$$A_2 = \frac{V_o}{V_{d1}} = -g_{m2}[R_{D2} \parallel (R_1 + R_2)]$$

$$A = \frac{V_o}{V_i} = A_1 A_2 = \frac{g_{m1}R_{D1}g_{m2}[R_{D2} \parallel (R_1 + R_2)]}{1 + g_{m1}(R_1 \parallel R_2)}$$

For the numerical values given,

$$A = \frac{4 \times 10 \times 4[10 \parallel (1+9)]}{1 + 4(1 \parallel 9)} = 173.9 \text{ V/V}$$

which is reasonably close to the value of 166.7 V/V we obtained in Example 11.3.

The value of β is determined from the β circuit in Fig. 11.18(c),

$$\beta \equiv \frac{V_f}{V_o} = \frac{R_1}{R_1 + R_2} = \frac{1}{1+9} = 0.1$$

which is identical to the value found in Example 11.3. The closed-loop gain V_o/V_s can now be found:

$$\frac{V_o}{V_s} = A_f = \frac{A}{1 + A\beta} = \frac{173.9}{1 + 173.9 \times 0.1} = 9.46 \text{ V/V}$$

which is very close to the value of 9.43 V/V obtained in Example 11.3.

The input resistance is obviously infinite because of the infinite input resistance of the MOSFET. The output resistance R_{out} is obtained as follows,

$$R_{out} = R_{of} = \frac{R_o}{1 + A\beta}$$

Example 11.6 *continued*

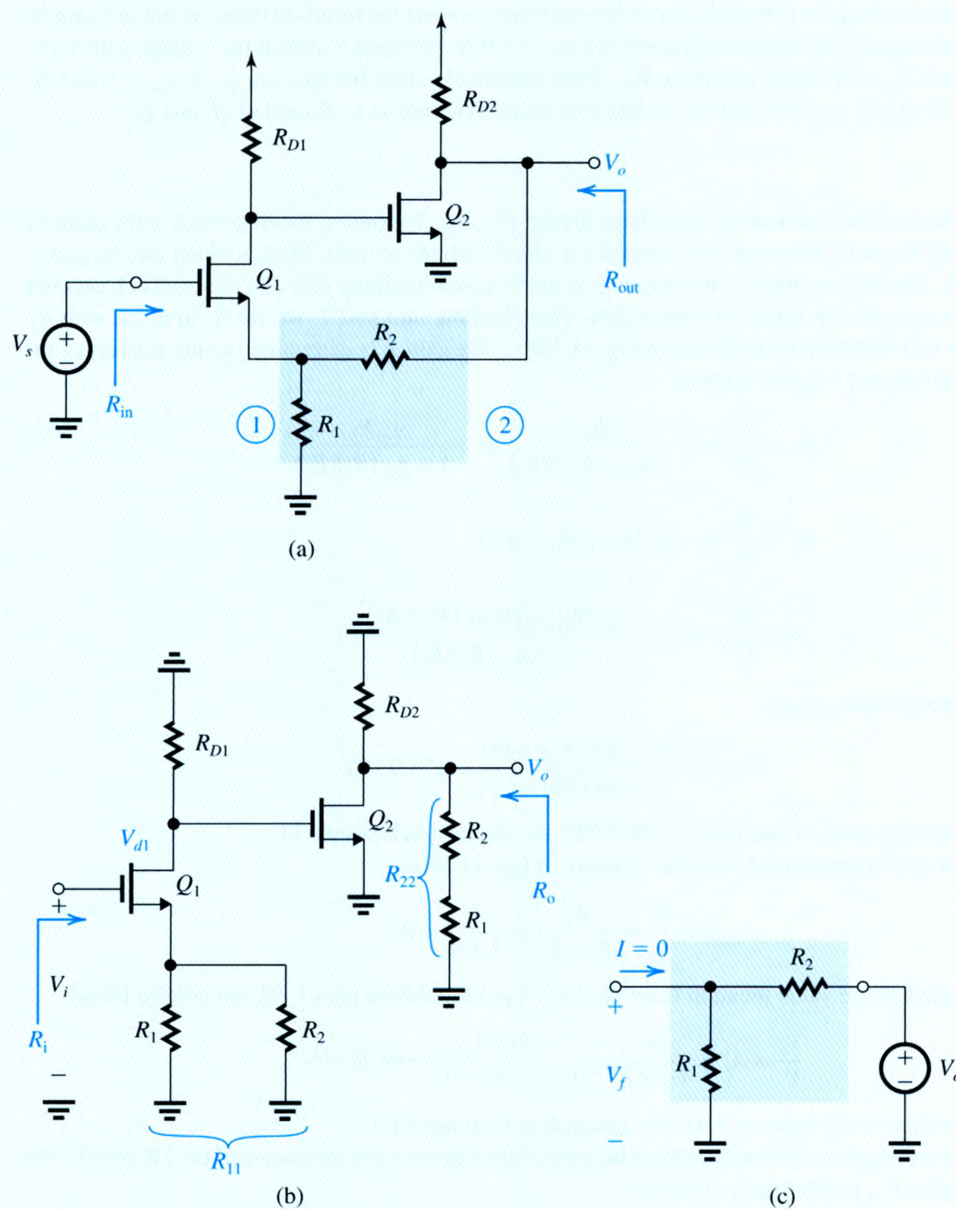


Figure 11.18 **(a)** Series–shunt feedback amplifier for Example 11.6; **(b)** the A circuit; **(c)** the β circuit.

where R_o is the output resistance of the A circuit. From Fig. 11.18(b),

$$R_o = R_{D2} \parallel (R_1 + R_2) = 10 \parallel 10 = 5\,\text{k}\Omega$$

The amount of feedback is

$$1 + A\beta = 1 + (173.9 \times 0.1) = 18.39$$

Thus,

$$R_{\text{out}} = \frac{5000}{18.39} = 272\ \Omega$$

which is relatively low given that the open-loop amplifier has $R_o = 5000\ \Omega$. Note that the loop-gain method used in Example 11.3 cannot provide the input and output resistances because the A circuit is not determined.

Video Example VE 11.2 Design and Analysis of a Feedback Voltage Amplifier

This problem deals with the series–shunt feedback amplifier of Fig. VE11.2. The current-mirror-loaded differential amplifier has a feedback network consisting of the voltage divider (R_1, R_2), with $R_1 + R_2 = 1$ MΩ. The devices are sized to operate at $|V_{OV}| = 0.2$ V. For all devices $|V_A| = 5$ V. The input signal source has a zero dc component.

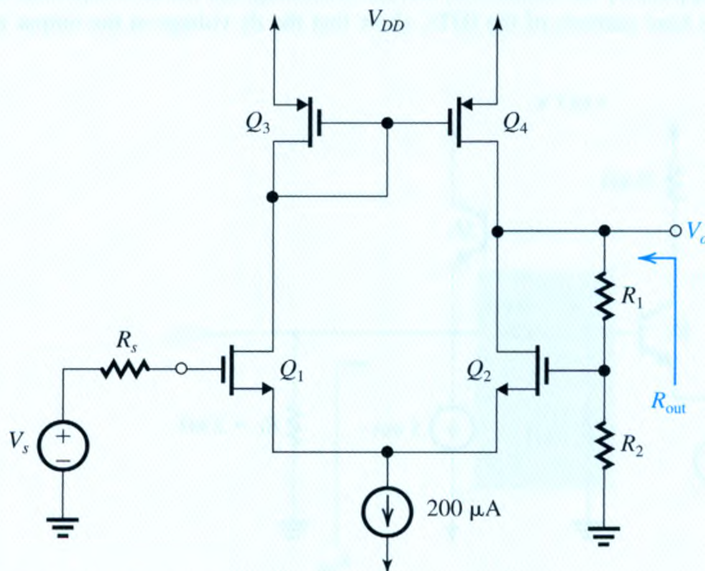


Figure VE11.2 Circuit for Video Example 11.2.

(a) Show that the feedback is negative.
(b) What do you expect the dc voltage at the gate of Q_2 to be? At the output? (Neglect the Early effect.)
(c) Find the A circuit. Derive an expression for A and find its value.
(d) Select values for R_1 and R_2 to obtain a closed-loop voltage gain $V_o/V_s = 5$ V/V.
(e) Find the value of R_{out}.
(f) Utilizing the open-circuit, closed-loop gain (5 V/V) and the value of R_{out} found in (e), find the value of gain obtained when a resistance $R_L = 10\ \text{k}\Omega$ is connected to the output.
(g) As an alternative approach to (f) above, redo the analysis of the A circuit including R_L. Then utilize the values of R_1 and R_2 found in (d) to determine β and A_f. Compare the value of A_f to that found in (f).


VE 11.2

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 11.40

EXERCISES

11.6 If the op amp of Example 11.5 has a uniform –20-dB/decade high-frequency rolloff with $f_{3dB} = 1$ kHz, find the 3-dB frequency of the closed-loop gain V_o/V_s.
Ans. 60.1 kHz

11.7 The circuit shown in Fig. E11.17 consists of a differential stage followed by an emitter follower, with series–shunt feedback supplied by the resistors R_1 and R_2. Assuming that the dc component of V_s is zero, and neglecting the base currents of the BJTs, show that the dc voltage at the output is

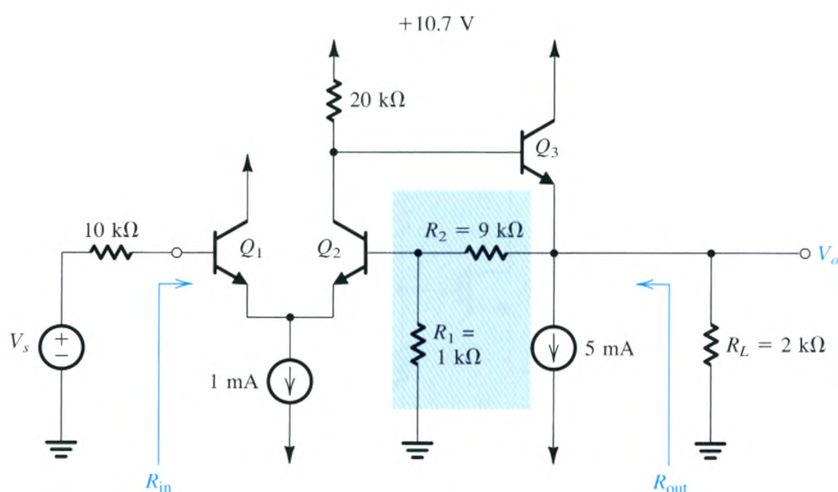


Figure E11.7

approximately zero and find the dc operating current of each of the three transistors. Then find the values of A, β, $A_f \equiv V_o/V_s$, R_{in}, and R_{out}. Assume that the transistors have $\beta = 100$.
Ans. 0.5 mA, 0.5 mA, 5 mA; 85 V/V; 0.1 V/V; 8.95 V/V; 189.5 kΩ; 19.2 Ω.

11.8 For the series–shunt amplifier in Fig. 11.9(c), which was considered in Exercise 11.5, find A, β, A_f, R_{in}, and R_{out}. Neglect r_o of Q. Compare results to those obtained in Exercise 11.5.
Ans. $A = g_m[R_D \| (R_1 + R_2)]$; $\beta = R_1/(R_1 + R_2)$;
$A_f = A/(1 + A\beta)$; $R_{in} = (1/g_m)(1 + A\beta)$;
$R_{out} = [R_D \| (R_1 + R_2)]/(1 + A\beta)$

Comparison: A and β are identical to the corresponding expressions found in Exercise 11.5. However, R_{in} and R_{out} cannot be determined using the method of Exercise 11.5.

11.5 Other Feedback-Amplifier Types

Having studied in detail the most common feedback-amplifier type, the feedback voltage amplifier, we now consider the three other types of feedback amplifier: the feedback transconductance amplifier, the feedback current amplifier, and the feedback transresistance amplifier. We will build on the study of the feedback voltage amplifier and give the results without derivation. We will illustrate the analysis method with a large number of worked-out examples dealing with practical and widely used circuits.

11.5.1 Basic Principles

1. *Sensing:* The feedback network must sample the output signal of interest. Thus if V_o is the output signal of interest, as in the case of voltage and transresistance amplifiers, the feedback network is connected in parallel (or *shunt*) with the amplifier output nodes, just as a voltmeter is connected to measure a voltage [see Fig. 11.19(a)]. On the other hand, if I_o is the output signal of interest, as in the case of transconductance and current amplifiers, the feedback network is connected in *series* with the output loop, just as a current meter is connected to measure a current [see Fig. 11.19(b)].
2. *Mixing:* If the input signal to be amplified is a voltage, as in the case of voltage and transconductance amplifiers, the signal source is represented by its Thévenin equivalent and the feedback voltage signal V_f is connected in *series* with the input signal source V_s [see Fig. 11.20(a)]. On the other hand, if the signal to be amplified is a current, as in the case of current and transresistance amplifiers, the Norton form is used to represent the signal source, and the feedback current signal I_f is connected in parallel (*shunt*) with the input signal source I_s [see Fig. 11.20(b)].
3. *Feedback topology:* From the above, it follows that for each of the four amplifier types there is a uniquely appropriate feedback topology:

Amplifier Type	Appropriate Feedback Topology	A_f
Voltage	Series–Shunt	V_o/V_s
Transconductance	Series–Series	I_o/V_s
Current	Shunt–Series	I_o/I_s
Transresistance	Shunt–Shunt	V_o/I_s

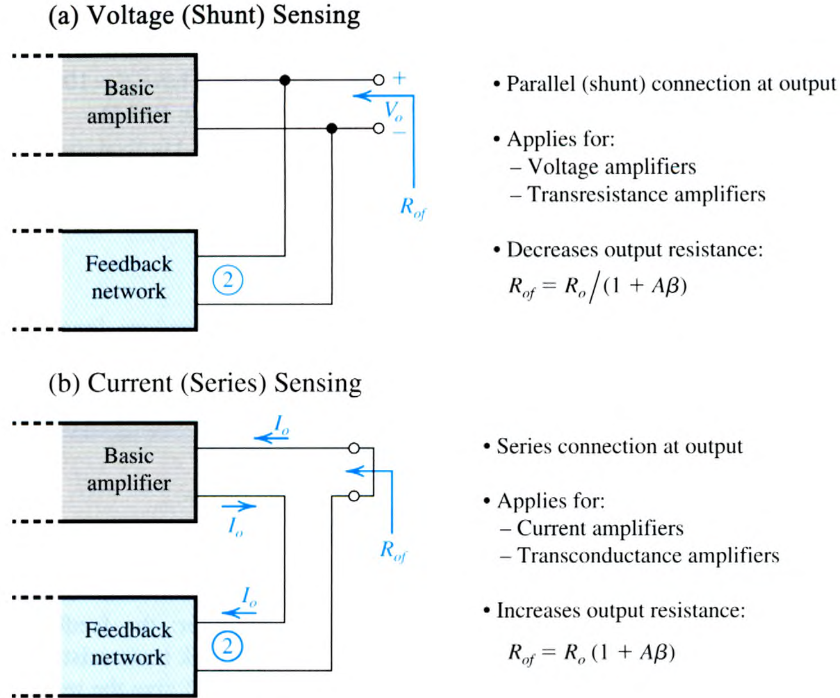
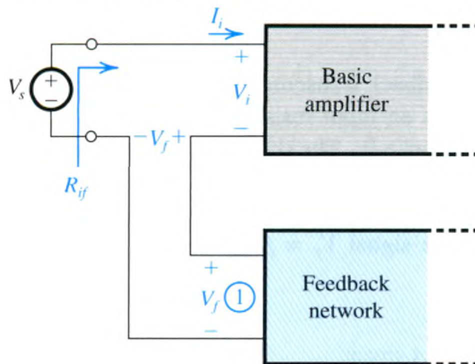


Figure 11.19 The two methods of sensing.

The appropriate feedback topology not only stabilizes the gain of interest (e.g., the transconductance $A_f \equiv I_o/V_s$ in a transconductance amplifier), but also makes the input and output resistances more ideal (e.g., the shunt–series topology decreases the input resistance and increases the output resistance of a current amplifier).

4. *Input and output resistance:* The increase or decrease of the input or output resistance depends *solely* on the type of connection: Series connection always increases the resistance; parallel (shunt) connection always decreases the resistance. Furthermore, the increase or decrease is always by the amount of feedback, $(1 + A\beta)$. Thus, as an example, for the feedback current amplifier, the shunt connection at the input decreases the input resistance; $R_{if} = R_i/(1 + A\beta)$, and the series connection at the output increases the output resistance; $R_{of} = (1 + A\beta)R_o$, where R_i and R_o are the input and output resistances of the open-loop amplifier (A circuit).
5. *Dimensions of $A, \beta, A\beta$, and A_f:* Depending on the amplifier type, A, β, and A_f have the dimensions of V/V, A/A, V/A, or A/V. However, $A\beta$ is always dimensionless. For a feedback transconductance amplifier, for example, $A \equiv I_o/V_i$ (A/V), $\beta \equiv V_f/I_o$ (V/A), $A_f \equiv I_o/V_s$ (A/V), and $A\beta$ is in V/V or essentially dimensionless.
6. *Determining β and $A_f|_{\text{ideal}}$:* To determine β, we set $A = \infty$. For the voltage (series) mixing case, this results in $V_i = 0$ and hence $I_i = 0$. From Fig. 11.20(a) we see that port 1 of the feedback network will be open-circuited and β is found as the ratio of the open-circuit voltage V_f to the output quantity being sensed. For the case of current (shunt) mixing, $A = \infty$ leads to $I_i = 0$ and correspondingly $V_i = 0$. From Fig. 11.20(b) we see that port 1 of the feedback network will be short-circuited and β is found as the ratio of the short-circuit current I_f to the output quantity being sensed. The ideal value of the closed-loop gain is found as $A_f|_{\text{ideal}} = 1/\beta$.

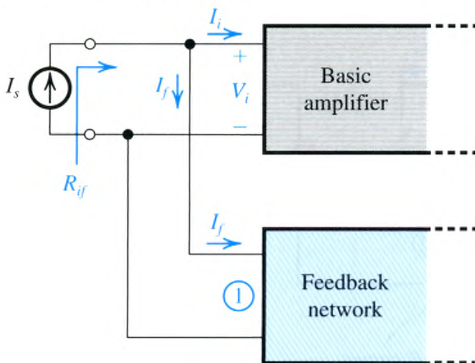
(a) Voltage (Series) Mixing



- Series connection at input
- Applies for:
  - Voltage amplifiers
  - Transconductance amplifiers
- Increases input resistance:

  $R_{if} = R_i\,(1 + A\beta)$

(b) Current (Shunt) Mixing



- Parallel (shunt) connection at input
- Applies for:
  - Current amplifiers
  - Transresistance amplifiers
- Decreases input resistance:

  $R_{if} = R_i \big/ (1 + A\beta)$

Figure 11.20 The two methods of mixing.

7. *Analysis using the loop gain:* For any feedback-amplifier type, we can find the loop gain $A\beta$ using the method described in Section 11.3.3. We can then use the loop gain together with the feedback factor β to determine the open-loop gain A and hence the closed-loop gain A_f. This approach, however, does not enable us to determine the input and output resistances of the feedback amplifier. For these, we need to obtain the A circuit using the systematic analysis approach described below.

EXERCISE

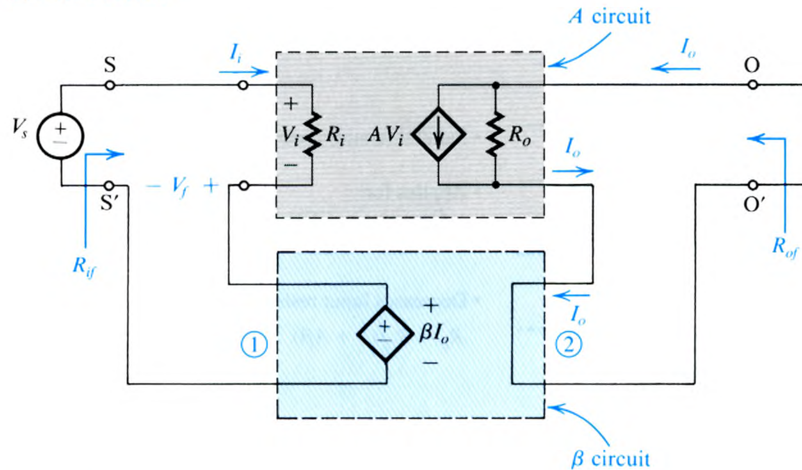
D 11.9 A feedback current amplifier is to have an ideal closed-loop gain of 10 A/A. What is the required value of the feedback factor β? If the open-loop amplifier has a gain of 1000 A/A, an input resistance of 1 kΩ, and an output resistance of 100 kΩ, find the realized values of closed-loop gain, input resistance, and output resistance.

Ans. 0.1 A/A; 9.9 A/A; 9.9 Ω; 10.1 MΩ

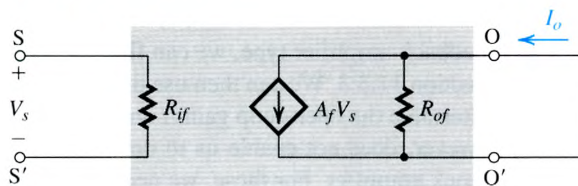
11.5.2 The Feedback Transconductance Amplifier (Series–Series)

Figure 11.21(a) shows the ideal structure of the feedback transconductance amplifier. The open-loop amplifier (A circuit) is unilateral and has an input resistance R_i, a short-circuit transconductance gain A in A/V, and an output resistance R_o. The short-circuit output current $I_o = AV_i$ is sensed by the feedback network. Note that since the resistance looking into port 2 of the feedback network is zero, the feedback network does not load the amplifier output. The feedback network provides at port 1 a voltage signal $V_f = \beta I_o$, where the feedback

(a) Ideal Structure



(b) Equivalent Circuit



(c) Formulas

$$A_f = \frac{I_o}{V_s} = \frac{A}{1 + A\beta}$$

$$A_f\big|_{\text{ideal}} = \frac{1}{\beta}$$

$$R_{if} = (1 + A\beta)R_i$$

$$R_{of} = (1 + A\beta)R_o$$

Figure 11.21 The feedback transconductance amplifier (series–series): **(a)** ideal structure; **(b)** equivalent circuit; **(c)** applicable formulas.

(a) General Structure

(b) Finding the A Circuit and β

(i) The A circuit is

where R_{11} is obtained from

and R_{22} is obtained from

and the gain A is defined $A \equiv \dfrac{I_o}{V_i}$

(ii) β is obtained from

$$\beta \equiv \left.\frac{V_f}{I_o}\right|_{I_1 = 0}$$

(c) Gain, Input, and Output Resistance

- Use the formulas in Fig. 11.21 to find $A_f|_{\text{ideal}}$, A_f, R_{if}, and R_{of}.
- R_{in} and R_{out} can then be found from

$$R_{in} = R_{if} - R_s$$
$$R_{out} = R_{of} - R_L$$

Figure 11.22 Systematic analysis of the feedback transconductance amplifier (series–series).

factor β has the dimensions of V/A. The feedback signal is connected in series with the input signal source V_s, and the feedback network does not load the amplifier input. Finally, note the definitions of the input resistance with feedback, R_{if}, and the output resistance with feedback, R_{of}. The latter is the resistance found by looking into the output loop between any two nodes, such as O and O' while setting $V_s = 0$.

The equivalent circuit of the feedback transconductance amplifier is shown in Fig. 11.21(b). Note that the closed-loop gain A_f is the ratio of the short-circuit output current I_o and the input voltage V_s, thus it is the short-circuit transconductance of the feedback amplifier. Also, R_{of} is the resistance seen between any two nodes in the output loop, such as O and O', while V_s is set to zero. The formulas for A_f, R_{if}, and R_{of} are found in Fig. 11.21(c).

With the ideal case in hand, we can now consider the general or practical case of the feedback transconductance amplifier, shown in Fig. 11.22(a). To be able to apply feedback analysis to this circuit, we have to find the A circuit and β. These are shown in Fig. 11.22(b). The A circuit is found by augmenting the basic amplifier with R_s and R_L and the two resistances R_{11} and R_{22}, which represent the loading effect of the feedback network on the basic amplifier at the input and output, respectively. Figure 11.22(b) shows how to find R_{11} and R_{22}. Here, the series connection at both the input and the output means that the other port of the feedback network must be left open-circuited. This is also the case when β is determined.

Finally, Fig. 11.22(c) gives the formulas for determining the actual values of the input and output resistances, R_{in} and R_{out}, of the feedback amplifier from R_{if} and R_{of}. To see how these formulas come about, note from Fig. 11.22(a) that unlike R_{if}, R_{in} does not include R_s, and unlike R_{of}, R_{out} does not include R_L.

Example 11.7 Feedback Transconductance Amplifier Using an Op Amp and a MOSFET

Figure 11.23(a) shows a feedback transconductance amplifier using an op amp with an NMOS transistor. The feedback network consists of a resistor R_F that senses the output current I_o (recall that the drain and source currents of the MOSFET are equal) and provides a feedback voltage that is subtracted from V_s by means of the differencing action of the op-amp input. Notice that the feedback topology is series–series, which is uniquely appropriate for transconductance amplifiers.

The op amp has a gain $\mu = 1000$ V/V, a differential input resistance $R_{id} = 100$ kΩ, and an output resistance $r_{o1} = 1$ kΩ. The MOSFET has $g_m = 2$ mA/V and $r_{o2} = 20$ kΩ. The feedback resistance $R_F = 1$ kΩ.

(a) Find $A_f\,|_{\text{ideal}}$ and β.
(b) Find the A circuit and determine A, R_i and R_o.
(c) Find A_f, R_{in} and R_{out}.

Solution

(a) To determine $A_f\,|_{\text{ideal}}$ and β we set the open-loop gain to infinity. This results in a zero input voltage between the op-amp input terminal, which in turn results in a zero current in the input loop, as shown in Fig. 11.23(b). We see that the voltage across R_F will be equal to V_s:

$$I_o R_F = V_s$$

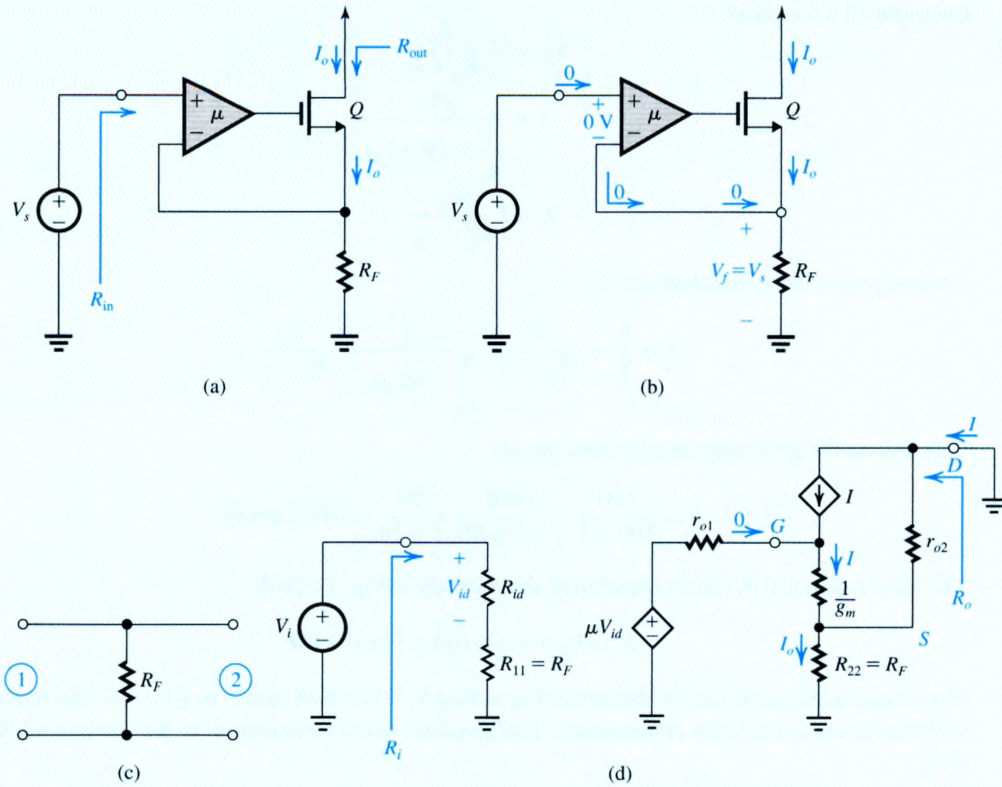


Figure 11.23 **(a)** Feedback transconductance amplifier for Example 11.7; **(b)** determining $A_f|_{ideal}$ by setting the open-loop gain to infinity; **(c)** the feedback network; **(d)** the A circuit with the op amp and the MOSFET replaced with their equivalent circuit models.

which results in

$$A_f|_{ideal} \equiv \frac{I_o}{V_s} = 1/R_F$$

and correspondingly

$$\beta = \frac{1}{A_f|_{ideal}} = R_F$$

For $R_F = 1\ \text{k}\Omega$,

$$A_f|_{ideal} = 1\ \text{mA/V} \qquad \beta = 1\ \text{V/mA}$$

(b) The feedback circuit is shown in Fig. 11.23(c), from which we can find the loading resistances:

$$R_{11} = R_F = 1\ \text{k}\Omega \qquad R_{22} = R_F = 1\ \text{k}\Omega$$

The A circuit is shown in Fig. 11.23(d), where the op amp and the MOSFET have been replaced with their equivalent circuit models. The open-loop gain A can be found by following the signal from input to output:

Example 11.7 *continued*

$$V_{id} = V_i \frac{R_{id}}{R_{id} + R_F}$$

$$I = \frac{\mu V_{id}}{\frac{1}{g_m} + (R_F \| r_{o2})}$$

$$I_o = I \frac{r_{o2}}{R_F + r_{o2}}$$

Combing these three equations gives

$$A \equiv \frac{I_o}{V_i} = \frac{R_{id}}{R_{id} + R_F} \frac{\mu}{\frac{1}{g_m} + (R_F \| r_{o2})} \frac{r_{o2}}{R_F + r_{o2}}$$

Substituting the given numerical values, we get

$$A = \frac{100}{100 + 1} \frac{1000}{\frac{1}{2} + (1 \| 20)} \frac{20}{1 + 20} = 649.2 \text{ mA/V}$$

The input resistance R_i can be found from the A circuit in Fig. 11.23(d):

$$R_i = R_{id} + R_F = 100 + 1 = 101 \text{ k}\Omega$$

The output resistance R_o can be determined by setting $V_i = 0$, which results in $\mu V_{id} = 0$. The transistor will then be operating in the common-gate configuration with a resistance R_F in the source lead. Thus, R_o is

$$R_o = r_{o2} + R_F + g_{m2} r_{o2} R_F$$
$$= 20 + 1 + 2 \times 20 \times 1 = 61 \text{ k}\Omega$$

(c) The loop gain $A\beta$ is

$$A\beta = 649.2 \times 1 = 649.2$$

and the amount of feedback is

$$1 + A\beta = 1 + 649.2 = 650.2$$

Now we can find the closed-loop transconductance A_f:

$$A_f \equiv \frac{I_o}{V_s} = \frac{A}{1 + A\beta} = \frac{649.2}{650.2} = 0.998 \text{ mA/V}$$

which is nearly equal to the ideal value of 1 mA/V. The input resistance is

$$R_{in} = R_{if} = (1 + A\beta) R_i = 650.2 \times 101 = 65.7 \text{ M}\Omega$$

and the output resistance is

$$R_{out} = R_{of} = (1 + A\beta) R_o = 650.2 \times 61 = 39.7 \text{ M}\Omega$$

EXERCISES

11.10 If in the circuit of Example 11.7, the gain μ is reduced from 1000 V/V to 100 V/V, find the new values of A, A_f, R_{in}, and R_{out}.
Ans. 64.9 mA/V; 0.985 mA/V; 6.7 MΩ; 4 MΩ

D11.11 If we want to double the ideal A_f of the feedback transconductance amplifier of Example 11.7, what value should R_F be changed to? What would A and A_f become?
Ans. 500 Ω; 982.7 mA/V; 1.996 mA/V.

Example 11.8 Feedback Transconductance Amplifier Using BJTs

Because negative feedback extends the amplifier bandwidth, it is commonly used in the design of wideband amplifiers. One such amplifier is the MC1553. Part of the circuit of the MC1553 is shown in Fig. 11.24(a). The circuit shown (called a **feedback triple**) is composed of three gain stages with series–series feedback provided by the network composed of R_{E1}, R_F, and R_{E2}.

Observe that the feedback network samples the emitter current I_o of Q_3, and thus I_o is the output quantity of the feedback amplifier. However, I_o is, practically, hard to use. Thus I_c, the collector current of Q_3, is commonly taken as the output. This current is of course almost equal to I_o; $I_c = \alpha I_o$. Thus, as a transconductance amplifier with I_c as the output current, the output resistance of interest is that labeled R_{out} in Fig. 11.24(a). In some applications, I_c is passed through a load resistance, such as R_{C3}, and the voltage V_o is taken as the output. Assume that the bias circuit, which is not shown, establishes $I_{C1} = 0.6$ mA, $I_{C2} = 1$ mA, and $I_{C3} = 4$ mA. Also assume that for all three transistors, $h_{fe} = 100$ and $r_o = \infty$.

(a) Find an expression and value for the ideal closed-loop gain $A_f \equiv I_o/V_s$ and hence for $\beta, I_c/V_s$. and V_o/V_s.
(b) Use feedback analysis to find A, A_f, V_o/V_s, R_{in}, and R_{out}. For the calculation of R_{out}, assume that r_o of Q_3 is 25 kΩ.

Solution

(a) To obtain $A_f\,|_{ideal}$ we set $A = \infty$. This results in the situation depicted in Fig. 11.24(b), where $V_i = 0$ and correspondingly $I_{e1} = 0$. We see that $V_f = V_s$. Analysis of this circuit to obtain $A_f\,|_{ideal} = I_o/V_s$ proceeds as follows: The current through the series connection of R_F and R_{E1} is V_f/R_{E1}. Thus, the voltage V_{e3} will be

$$V_{e3} = \frac{V_f}{R_{E1}}(R_{E1} + R_F)$$

Example 11.8 *continued*

and a node equation at the emitter of Q_3 provides

$$I_o = \frac{V_f}{R_{E1}} + \frac{V_f}{R_{E1}R_{E2}}(R_{E1} + R_F)$$

Substituting $V_f = V_s$ and rearranging gives the ideal value of the closed-loop gain:

$$A_f \,|_{\text{ideal}} \equiv \frac{I_o}{V_s} = \frac{1}{R_{E1}}\left(1 + \frac{R_{E1}}{R_{E2}} + \frac{R_F}{R_{E2}}\right)$$

and the feedback factor β is

$$\beta = \frac{1}{A_f \,|_{\text{ideal}}} = R_{E1}\Big/\left(1 + \frac{R_{E1}}{R_{E2}} + \frac{R_F}{R_{E2}}\right)$$

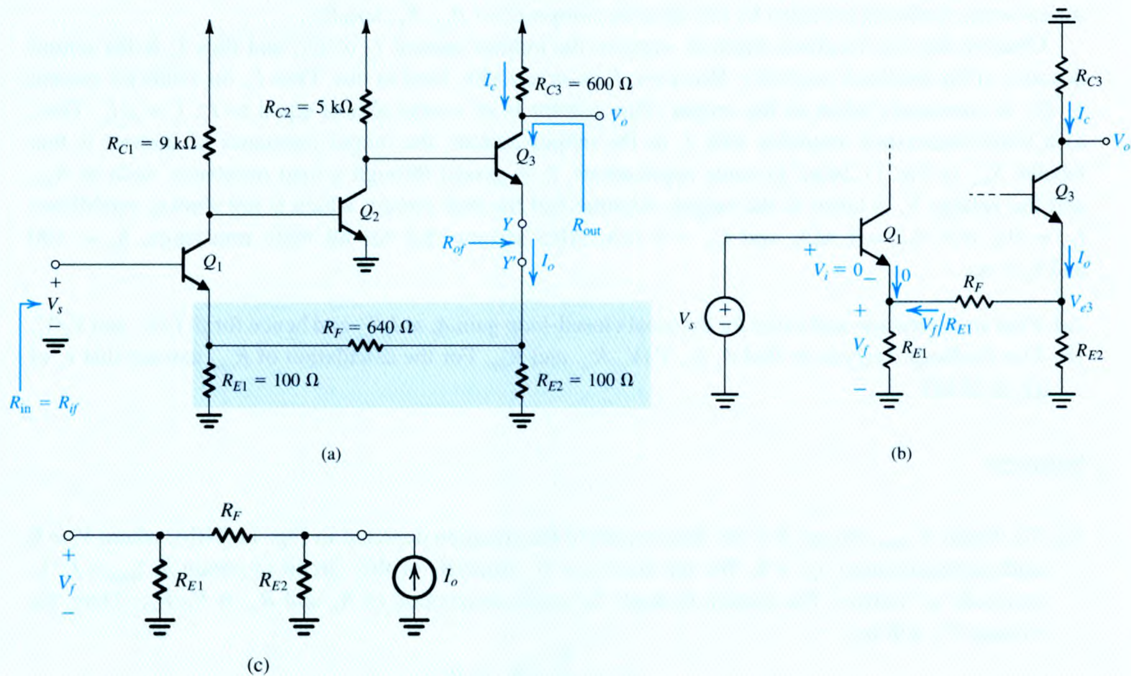


Figure 11.24 Circuits for Example 11.8.

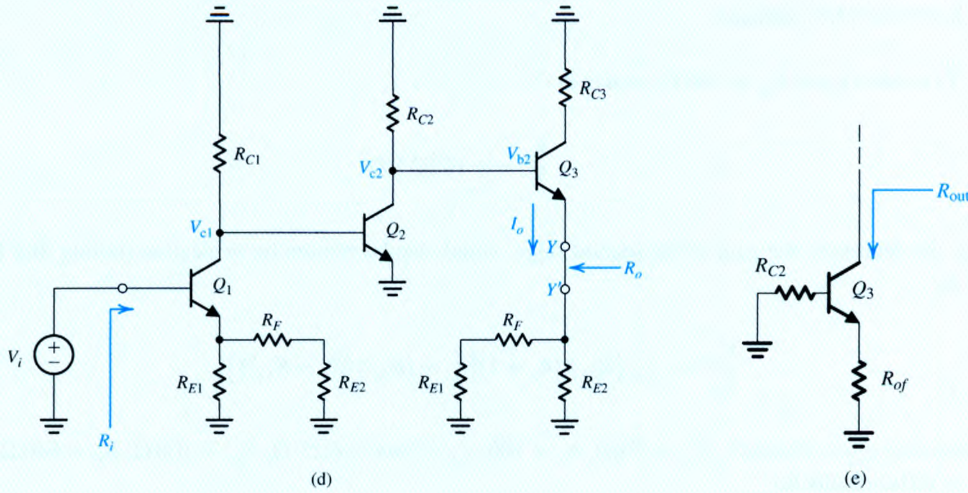


Figure 11.24 *continued*

Substituting the numerical values given for R_{E1}, R_{E2}, and R_F, we obtain

$$A_f\,|_{\text{ideal}} = 84 \text{ mA/V}$$

and

$$\beta = 0.0119 \text{ V/mA}$$

Note that β could also be found from the feedback circuit in Fig. 11.24(c) as V_f/I_o.

Finally, we find

$$\frac{I_c}{V_s} \simeq \frac{I_o}{V_s} = 84 \text{ mA/V}$$

$$\frac{V_o}{V_s} = \frac{-I_c R_{C3}}{V_s} = -84 \times 0.6 = -50.4 \text{ V/V}$$

(b) Using the loading rules given in Fig. 11.22, we obtain the A circuit shown in Fig. 11.24(d). To find $A \equiv I_o/V_i$ we first determine the gain of the first stage. This can be written by inspection as

$$\frac{V_{c1}}{V_i} = \frac{-\alpha_1 (R_{C1} \,\|\, r_{\pi 2})}{r_{e1} + [R_{E1} \,\|\, (R_F + R_{E2})]}$$

Since Q_1 is biased at 0.6 mA, $r_{e1} = 41.7\ \Omega$. Transistor Q_2 is biased at 1 mA; thus $r_{\pi 2} = h_{fe}/g_{m2} = 100/40 = 2.5\ \text{k}\Omega$. Substituting these values together with $\alpha_1 = 0.99$, $R_{C1} = 9\ \text{k}\Omega$, $R_{E1} = 100\ \Omega$,

Example 11.8 *continued*

$R_F = 640\,\Omega$, and $R_{E2} = 100\,\Omega$, results in

$$\frac{V_{c1}}{V_i} = -14.92 \text{ V/V}$$

Next, we determine the gain of the second stage, which can be written by inspection (noting that $V_{b2} = V_{c1}$) as

$$\frac{V_{c2}}{V_{c1}} = -g_{m2}\{R_{C2} \| (h_{fe}+1)[r_{e3} + (R_{E2} \| (R_F + R_{E1}))]\}$$

Substituting $g_{m2} = 40\,\text{mA/V}$, $R_{C2} = 5\text{ k}\Omega$, $h_{fe} = 100$, $r_{e3} = 25/4 = 6.25\ \Omega$, $R_{E2} = 100\,\Omega$, $R_F = 640\,\Omega$, and $R_{E1} = 100\,\Omega$ results in

$$\frac{V_{c2}}{V_{c1}} = -131.2 \text{ V/V}$$

Finally, for the third stage we can write by inspection

$$\frac{I_o}{V_{c2}} = \frac{I_{e3}}{V_{b3}} = \frac{1}{r_{e3} + (R_{E2} \| (R_F + R_{E1}))}$$

$$= \frac{1}{6.25 + (100 \| 740)} = 10.6 \text{ mA/V}$$

Combining the gains of the three stages results in

$$A \equiv \frac{I_o}{V_i} = -14.92 \times -131.2 \times 10.6 \times 10^{-3}$$

$$= 20.7 \text{ A/V}$$

We can now find the closed-loop gain A_f from

$$A_f \equiv \frac{I_o}{V_s} = \frac{A}{1 + A\beta}$$

$$= \frac{20.7}{1 + 20.7 \times 11.9} = 83.7 \text{ mA/V}$$

which we note is very close to the ideal value found in (a) above. This is not surprising, since the loop gain $A\beta = 20.7 \times 11.9 = 246.3$ is large ($\gg 1$).

The voltage gain is found from

$$\frac{V_o}{V_s} = \frac{-I_c R_{C3}}{V_s} \simeq \frac{-I_o R_{C3}}{V_s} = -A_f R_{C3}$$
$$= -83.7 \times 10^{-3} \times 600 = -50.2 \text{ V/V}$$

which is also very close to the ideal value found in (a) above.

The input resistance of the feedback amplifier is given by

$$R_{\text{in}} = R_{if} = R_i(1 + A\beta)$$

where R_i is the input resistance of the A circuit. The value of R_i can be found from the circuit in Fig. 11.24(d) as follows:

$$R_i = (h_{fe}+1)[r_{e1} + (R_{E1} \,\|\, (R_F + R_{E2}))]$$
$$= 13.11 \text{ k}\Omega$$

Thus,

$$R_{if} = 13.11(1 + 20.7 \times 11.9) = 3.24 \text{M}\Omega$$

To find the output resistance R_{out}, which is the resistance looking into the collector of Q_3, we face a dilemma. The feedback does not sample I_c and thus we cannot use the feedback formulas directly. Nevertheless, we present a somewhat indirect solution to this problem below. Of course, we would not have this problem, had Q_1 been a MOSFET, since $I_d = I_s$ (see Example 11.7).

Since the feedback senses the emitter current I_o, the output resistance given by the feedback analysis will be the resistance seen in the emitter circuit, say between Y and Y',

$$R_{of} = R_o(1 + A\beta)$$

where R_o can be determined from the A circuit in Fig. 11.24(d) by breaking the circuit between Y and Y'. The resistance looking between these two nodes can be found to be

$$R_o = [R_{E2} \,\|\, (R_F + R_{E1})] + r_{e3} + \frac{R_{C2}}{h_{fe}+1}$$

which, for the values given, yields $R_o = 143.9\ \Omega$. The output resistance R_{of} of the feedback amplifier can now be found as

$$R_{of} = R_o(1 + A\beta) = 143.9(1 + 20.7 \times 11.9) = 35.6 \text{ k}\Omega$$

We can now use the value of R_{of} to obtain an approximate value for R_{out}. To do this, we assume that the effect of the feedback is to place a resistance R_{of} (35.6 kΩ) in the emitter of Q_3, and find the output resistance from the equivalent circuit shown in Fig. 11.24(e). This is the output resistance of a BJT with a resistance R_{of} in its emitter and a resistance R_{C2} in its base. Unfortunately, the formula we have for this (Eq. 8.66) does not account for a resistance in the base. The formula, however, can be modified (see

Example 11.8 *continued*

Problem 11.54) to obtain

$$R_{\text{out}} = r_{o3} + \left[R_{of} \,\|\, (r_{\pi 3} + R_{C2})\right]\left[1 + g_{m3} r_{o3} \frac{r_{\pi 3}}{r_{\pi 3} + R_{C2}}\right]$$

$$= 25 + [35.6 \,\|\, (0.625 + 5)]\left[1 + 160 \times 25 \times \frac{0.625}{0.625 + 5}\right]$$

$$= 2.19\,\text{M}\Omega$$

Thus R_{out} is increased (from r_{o3}) but not by $(1 + A\beta)$.

Video Example VE 11.3 Design and Systematic Analysis of a Feedback Transconductance Amplifier

The feedback transconductance amplifier in Fig. P11.45(a) uses a differential voltage amplifier with a gain μ and a very high input resistance, and a MOSFET Q having $g_m = 4$ mA/V and $r_o = 25$ kΩ. Design the circuit to obtain an ideal closed-loop transconductance of 5 mA/V and an output resistance R_{out} of 4 MΩ. What are the values required for R_F and μ? Also, what is the actual value obtained of the closed-loop transconductance?


VE 11.3

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 11.49

EXERCISES

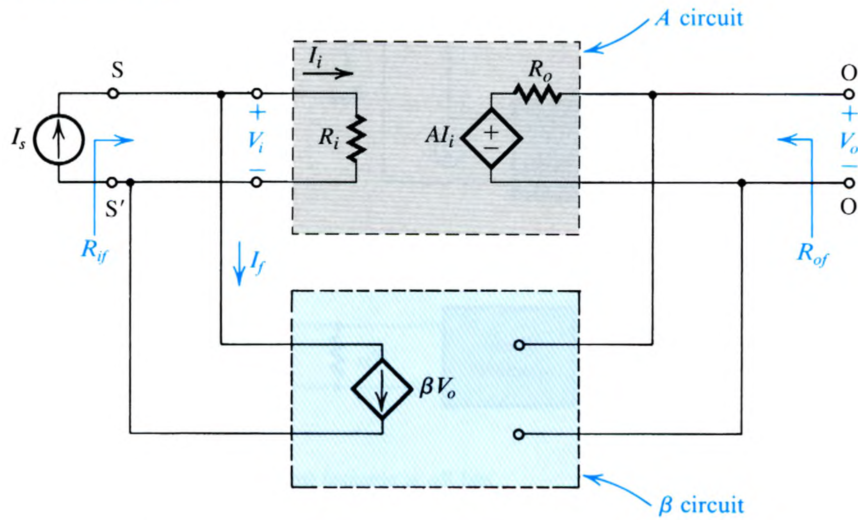
D11.12 For the feedback triple in Fig. 11.24(a), analyzed in Example 11.8, modify the value of R_F to obtain a closed-loop transconductance I_o/V_s of ideally 100 mA/V. What is the new value of R_F? For this value, what is the ideal value of the voltage gain if the output voltage is taken at the collector of Q_3?
Ans. 800 Ω; –60 V/V

11.13 Determine the loop gain of the feedback amplifier of Fig. 11.24(a). Set $V_s = 0$, break the loop between the collector of Q_1 and the base of Q_2, apply a voltage V_t to the base of Q_2, and connect a resistance equal to $r_{\pi 2}$ between the collector of Q_1 and ground. Find $A\beta$ as $(-V_r/V_t)$ where $V_r = V_{c1}$.
Ans. $A\beta = 249.2$ (slightly different from the value found in Example 11.8 because of the approximations inherent in the systematic feedback-analysis method).

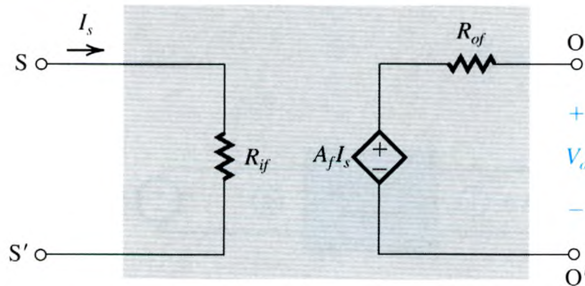
11.5.3 The Feedback Transresistance Amplifier (Shunt–Shunt)

Figure 11.25(a) shows the ideal feedback transresistance amplifier, which, as expected, uses the shunt–shunt topology. The amplifier equivalent circuit is shown in Fig. 11.25(b), with the formulas for determining A_f, R_{if}, and R_{of} given in Fig. 11.25(c).

(a) Ideal Structure



(b) Equivalent Circuit



(c) Formulas

$$A_f \equiv \frac{V_o}{I_s} = \frac{A}{1+A\beta}$$

$$A_f\Big|_{\text{ideal}} = \frac{1}{\beta}$$

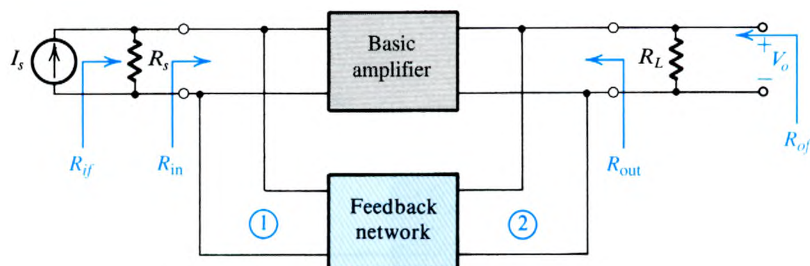
$$R_{if} = R_i/(1+A\beta)$$

$$R_{of} = R_o/(1+A\beta)$$

Figure 11.25 The feedback transresistance amplifier (shunt–shunt).

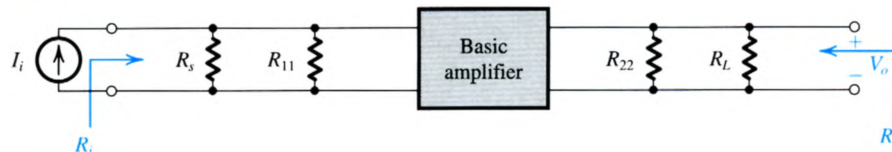
All the steps and formulas needed to apply the feedback-analysis method to a general transresistance amplifier are shown in Fig. 11.26. The method will be illustrated by a detailed example.

(a) General Structure

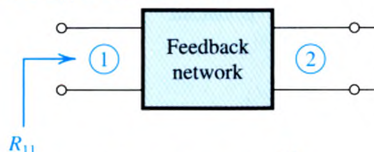


(b) Finding the A Circuit and β

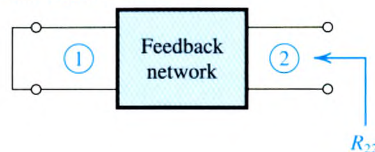
(i) The A circuit is



where R_{11} is obtained from

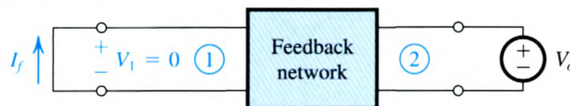


and R_{22} is obtained from



and the gain A is defined $A \equiv \dfrac{V_o}{I_i}$

(ii) β is obtained from



$$\beta \equiv \left.\frac{I_f}{V_o}\right|_{V_1 = 0}$$

(c) Gain, Input, and Output Resistance

- Use the formulas in Fig. 11.25 to find A_f, R_{if}, and R_{of}.
- R_{in} and R_{out} can then be found from

$$R_{in} = 1\Big/\left(\frac{1}{R_{if}} - \frac{1}{R_s}\right)$$

$$R_{out} = 1\Big/\left(\frac{1}{R_{of}} - \frac{1}{R_L}\right)$$

Figure 11.26 Analysis of the feedback transresistance amplifier (shunt–shunt).

Example 11.9 Feedback Transresistance Amplifier (Shunt–Shunt)

Figure 11.27(a) shows a feedback transresistance amplifier. It employs shunt–shunt feedback formed by connecting a resistance R_F in the negative-feedback path of a voltage amplifier with gain μ, an input resistance R_{id}, and an output resistance r_o. The amplifier μ can be implemented with an op amp, a simple differential amplifier, a single-ended inverting amplifier, or, in the limit, a single-transistor CE or CS amplifier. The latter case will be considered in Exercise 11.15. Of course, the higher the gain μ, the more ideal the characteristics of the feedback transresistance amplifier will be, simply because of the concomitant increase in loop gain. Use the following numerical values: $\mu = 10^4$ V/V, $R_{id} = 100$ kΩ, $r_o = 1$ kΩ, $R_s = R_L = 2$ kΩ, and $R_F = 10$ kΩ.

(a) Find the ideal closed-loop gain V_o/I_s and the feedback factor β.
(b) Find the A circuit and hence A, R_i, and R_o.
(c) Find A_f, R_{if}, R_{in}, R_{of}, and R_{out}.
(d) If instead of a current source I_s having a source resistance $R_s = 2$ kΩ, the amplifier is fed from a voltage source V_s having a source resistance $R_s = 2$ kΩ, find the voltage gain V_o/V_s.

Solution

(a) To obtain the ideal closed-loop gain A_f, we let the open-loop gain become infinite. From the circuit in Fig. 11.27(b) we see that $I_i = 0$, which results in $V_i = 0$. From the input node we see that

$$I_f = I_s$$

The output voltage V_o can be found as

$$V_o = V_i - I_f R_F = 0 - I_s R_F$$

Thus,

$$A_f\,|_{\text{ideal}} \equiv \frac{V_o}{I_s} = -R_F = -10\text{ k}\Omega = -10\text{ V/mA}$$

The feedback factor β can now be found as

$$\beta = \frac{1}{A_f\,|_{\text{ideal}}} = -0.1\text{ mA/V}$$

(b) The A circuit is shown in Fig. 11.27(d), where the loading effects of the feedback circuit, R_{11} and R_{22}, are obtained from the feedback circuit shown in Fig. 11.27(c) by short-circuiting ports 2 and 1, respectively. To determine A, we analyze the A circuit as follows:

$$V_{id} = I_i R_i$$

where

$$R_i = R_s \| R_{11} \| R_{id} = R_s \| R_F \| R_{id}$$
$$= 2 \| 10 \| 100 = 1.639\text{ k}\Omega$$
$$V_o = -\mu V_{id} \frac{R_L \| R_{22}}{(R_L \| R_{22}) + r_o}$$
$$= -\mu I_i R_i \frac{R_L \| R_F}{(R_L \| R_F) + r_o}$$

Example 11.9 *continued*

from which we can find

$$A \equiv \frac{V_o}{I_i} = -\mu R_i \frac{R_L \| R_F}{(R_L \| R_F) + r_o}$$
$$= -10^4 \times 1.639 \times \frac{(2\|10)}{(2\|10)+1} = -10,244 \text{ V/mA}$$

The open-loop output resistance R_o is obtained from the A circuit in Fig. 11.27(d) by setting $I_i = 0$, which results in $V_{id} = 0$ and thus

$$R_o = R_L \| R_{22} \| r_o = 2\|10\|1 = 625\ \Omega$$

(c)

$$A_f = \frac{A}{1+A\beta} = \frac{-10,244}{1+(-10,244 \times -0.1)}$$
$$= -\frac{10,244}{1+1024.4} = -9.990 \text{ V/mA}$$
$$R_{if} = \frac{R_i}{1+A\beta} = \frac{1.639 \text{ k}\Omega}{1025.4} = 1.6\ \Omega$$

Since $R_{if} = R_s \| R_{\text{in}}$, then

$$\frac{1}{R_{\text{in}}} = \frac{1}{R_{if}} - \frac{1}{R_s} = \frac{1}{1.6\ \Omega} - \frac{1}{2 \text{ k}\Omega}$$
$$\Rightarrow R_{\text{in}} \simeq 1.6\ \Omega$$
$$R_{of} = \frac{R_o}{1+A\beta} = \frac{625\ \Omega}{1025.4} = 0.61\ \Omega$$

Since $R_{of} = R_L \| R_{\text{out}}$, then

$$\frac{1}{R_{\text{out}}} = \frac{1}{R_{of}} - \frac{1}{R_L} = \frac{1}{0.61\ \Omega} - \frac{1}{2 \text{ k}\Omega}$$
$$\Rightarrow R_{\text{out}} \simeq 0.61\ \Omega$$

(d) If the amplifier is fed with a voltage source V_s having a resistance $R_s = 2$ kΩ, we can convert this source to its Norton's equivalent composed of $I_s = V_s/R_s$ and a parallel resistance R_s. This is the case we analyzed above, where we found $V_o = A_f I_s$. Thus,

$$V_o = A_f \frac{V_s}{R_s}$$
$$\Rightarrow \frac{V_o}{V_s} = \frac{A_f}{R_s} = \frac{-9.990 \text{ k}\Omega}{2 \text{ k}\Omega} = -4.995 \text{ V/V}$$

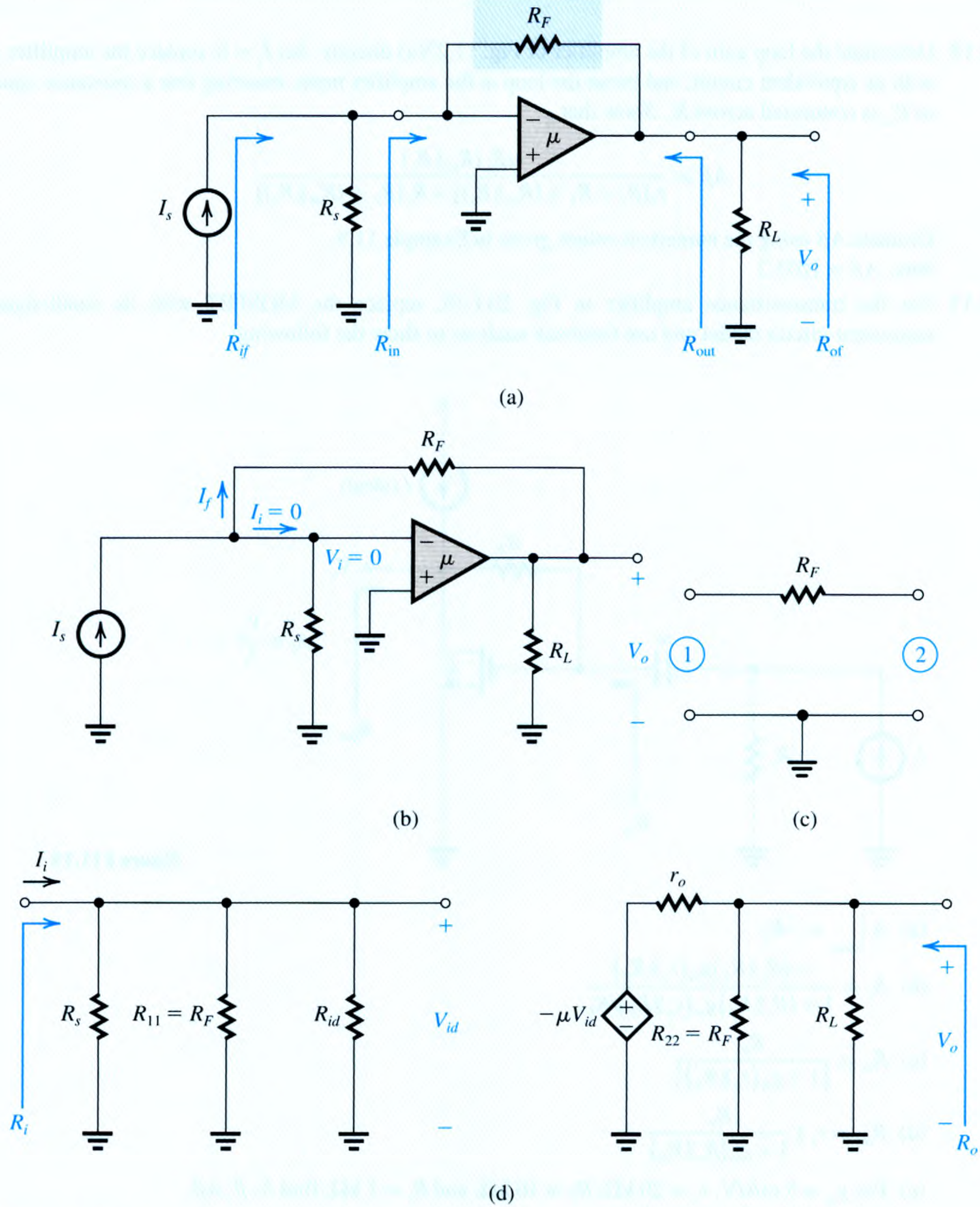


Figure 11.27 (**a**) A feedback transresistance amplifier; (**b**) determining $A_f|_{\text{ideal}}$ by setting $A = \infty$; (**c**) the feedback circuit; (**d**) the A circuit.

EXERCISES

11.14 Determine the loop gain of the amplifier of Fig. 11.27(a) directly. Set $I_s = 0$, replace the amplifier μ with its equivalent circuit, and break the loop at the amplifier input, ensuring that a resistance equal to R_{id} is connected across R_s. Show that

$$A\beta = \frac{\mu R_L (R_{id} \| R_s)}{r_o[R_L + R_F + (R_{id} \| R_s)] + R_L[R_F + (R_{id} \| R_s)]}$$

Evaluate $A\beta$ using the numerical values given in Example 11.9.

Ans. $A\beta = 1035.2$

11.15 For the transresistance amplifier in Fig. E11.15, replace the MOSFET with its small-signal equivalent-circuit model and use feedback analysis to show the following:

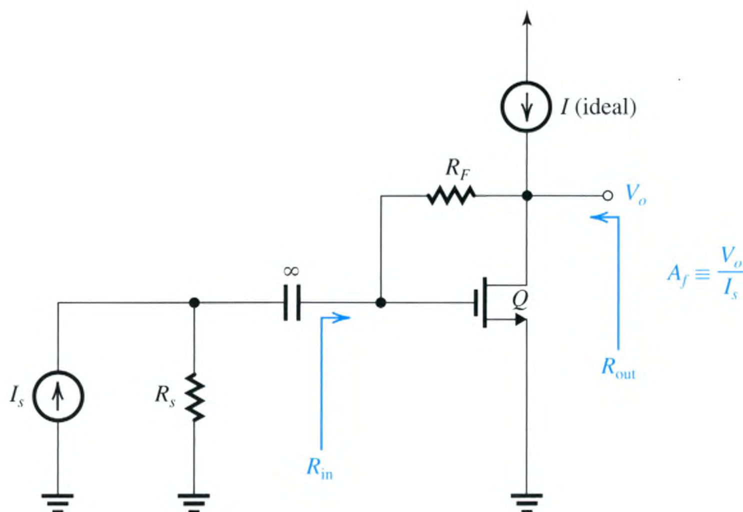


Figure E11.15

(a) $A_f\Big|_{\text{ideal}} = -R_F$.

(b) $A_f = \dfrac{-(R_s \| R_F) g_m (r_o \| R_F)}{1 + (R_s \| R_F) g_m (r_o \| R_F)/R_F}$

(c) $R_{in} = \dfrac{R_F}{[1 + g_m (r_o \| R_F)]}$

(d) $R_{out} = r_o \| \dfrac{R_F}{1 + g_m (R_s \| R_F)}$

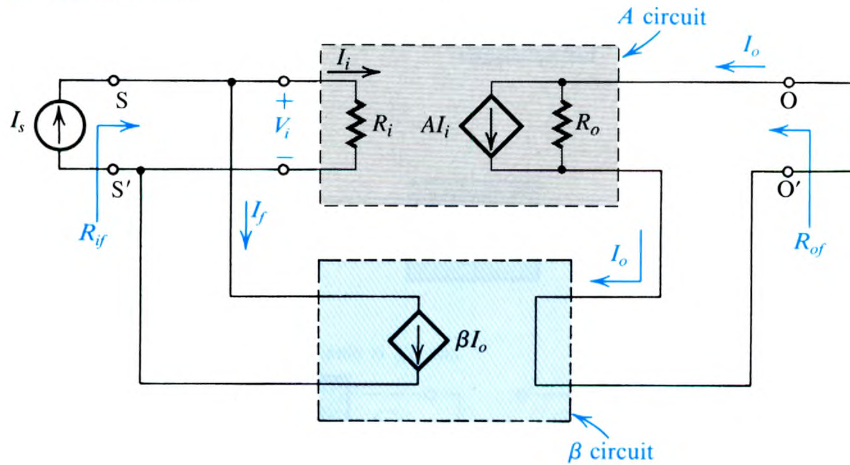
(e) For $g_m = 5$ mA/V, $r_o = 20$ kΩ, $R_F = 10$ kΩ, and $R_s = 1$ kΩ, find A, β, $A\beta$, A_f R_i, R_o, R_{if}, R_{in}, R_{of}, and R_{out}.

Ans. (e) −30.3 kΩ; −0.1 mA/V; 3.03; −7.52 kΩ (compare to the ideal value of −10 kΩ); 909 Ω; 6.67 kΩ; 226 Ω; 291 Ω; 1.66 kΩ; 1.66 kΩ

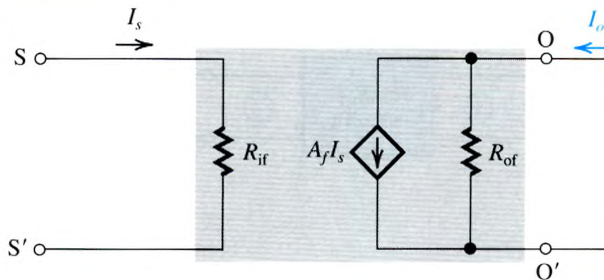
11.5.4 The Feedback Current Amplifier (Shunt–Series)

Figure 11.28(a) shows the ideal feedback current amplifier, which, as expected, utilizes the shunt–series topology. The amplifier equivalent circuit is shown in Fig. 11.28(b), with the formulas for determining A_f, $A_f\big|_{\text{ideal}}$, R_{if}, and R_{of} given in Fig. 11.28(c).

(a) Ideal Structure



(b) Equivalent Circuit



(c) Formulas

$$A_f \equiv \frac{I_o}{I_s} = \frac{A}{1+A\beta}$$

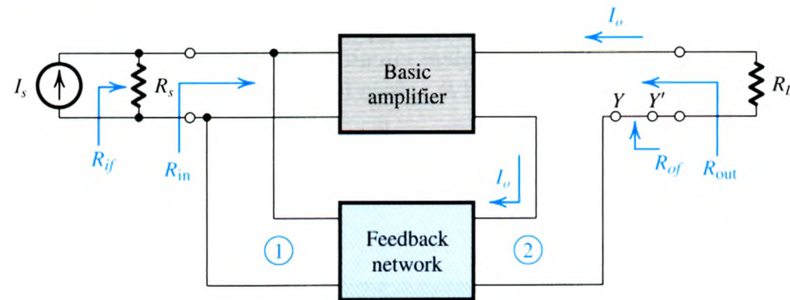
$$A_f\big|_{\text{ideal}} = \frac{1}{\beta}$$

$$R_{if} = R_i/(1+A\beta)$$

$$R_{of} = (1+A\beta)R_o$$

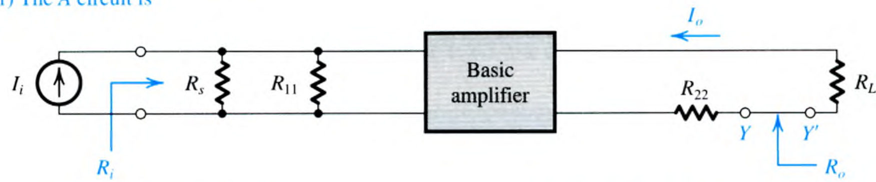
Figure 11.28 The feedback current amplifier (shunt–series).

(a) General Structure

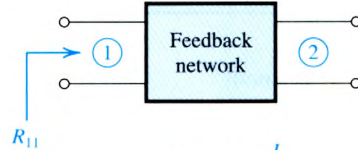


(b) Finding the A Circuit and β

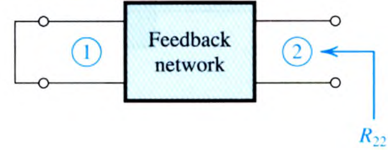
(i) The A circuit is



where R_{11} is obtained from

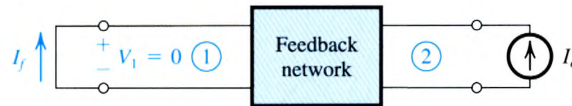


and R_{22} is obtained from



and the gain A is defined as $A \equiv \frac{I_o}{I_i}$

(ii) β is obtained from



$$\beta \equiv \left.\frac{I_f}{I_o}\right|_{V_1 = 0}$$

(c) Gain, Input, and Output Resistance

- Use the formulas in Fig. 11.28 to find A_f, R_{if}, and R_{of}.
- R_{in} and R_{out} can then be found from

$$R_{in} = 1 \Big/ \left(\frac{1}{R_{if}} - \frac{1}{R_s}\right)$$

$$R_{out} = R_{of} - R_L$$

Figure 11.29 Analysis of the feedback current amplifier (shunt–series).

Figure 11.29 shows how the feedback-analysis method can be applied to a general feedback current amplifier. All the analysis steps and formulas are given. We will now illustrate the method with a detailed example.

Example 11.10 Feedback Current Amplifier

Figure 11.30(a) shows a feedback current amplifier formed by cascading an inverting voltage amplifier μ with a MOSFET Q. The output current I_o is the drain current of Q. The feedback network, consisting of resistors R_1 and R_2, senses an exactly equal current, namely, the source current of Q, and provides a feedback current signal that is mixed with I_s at the input node. Note that the bias arrangement is *not* shown.

The amplifier μ can be implemented in a variety of ways, including by means of an op amp, a differential amplifier, or a single-ended inverting amplifier. The simplest approach is to implement μ with a CS MOSFET amplifier. However, in such a case the loop gain will be very limited. Assume that the amplifier μ has an input resistance R_{id}, an open-circuit voltage gain μ, and an output resistance r_{o1}.

Use the following numerical values: $\mu = 1000$ V/V, $R_{id} = \infty$, $r_{o1} = 1$ kΩ, $g_m = 5$ mA/V, $r_o\big|_Q = 20$ kΩ, $R_s = \infty$, $R_1 = 10$ kΩ, and $R_2 = 90$ kΩ.

(a) Find the ideal closed-loop gain I_o/I_s and the feedback factor β.
(b) Find the A circuit and hence A, R_i, and R_o.
(c) Find A_f, R_{if}, R_{in}, R_{of}, and R_{out}.

Solution

(a) To find $A_f\,|_{\text{ideal}}$ we let A become ∞. The resulting situation is illustrated in Fig. 11.30(b), where $I_i = 0$ and correspondingly $V_i = 0$. We can see that at the input node, $I_f = I_s$, and we can carry out the analysis shown in the figure to determine I_o,

$$I_o = -I_s\left(1 + \frac{R_2}{R_1}\right)$$

Thus,

$$A_f\,|_{\text{ideal}} \equiv \frac{I_o}{I_s} = -\left(1 + \frac{R_2}{R_1}\right)$$

For the given values of R_1 and R_2,

$$A_f\,|_{\text{ideal}} = -10 \text{ A/A}$$

We can now determine the value of β:

$$\beta = 1/A_f\,|_{\text{ideal}}$$

Thus,

$$\beta = -\frac{1}{1 + \frac{R_2}{R_1}} = -\frac{R_1}{R_1 + R_2} = -0.1$$

The two-port feedback circuit is shown in Fig.11.30(c). As an alternative approach, we could have used this circuit to determine β and then $A_f\,|_{\text{ideal}}$ as $(1/\beta)$.

(b) To obtain the A circuit we load the input side of the basic amplifier with R_s and R_{11}. The latter in this case is simply $R_1 + R_2$ (because port 2 of the feedback network is opened). We also load the output of the basic amplifier with R_{22}, which in this case is $R_1 \parallel R_2$ (because port 1 of the feedback network is shorted). The resulting A circuit is shown in Fig. 11.30(d), where we have replaced the amplifier μ

Example 11.10 *continued*

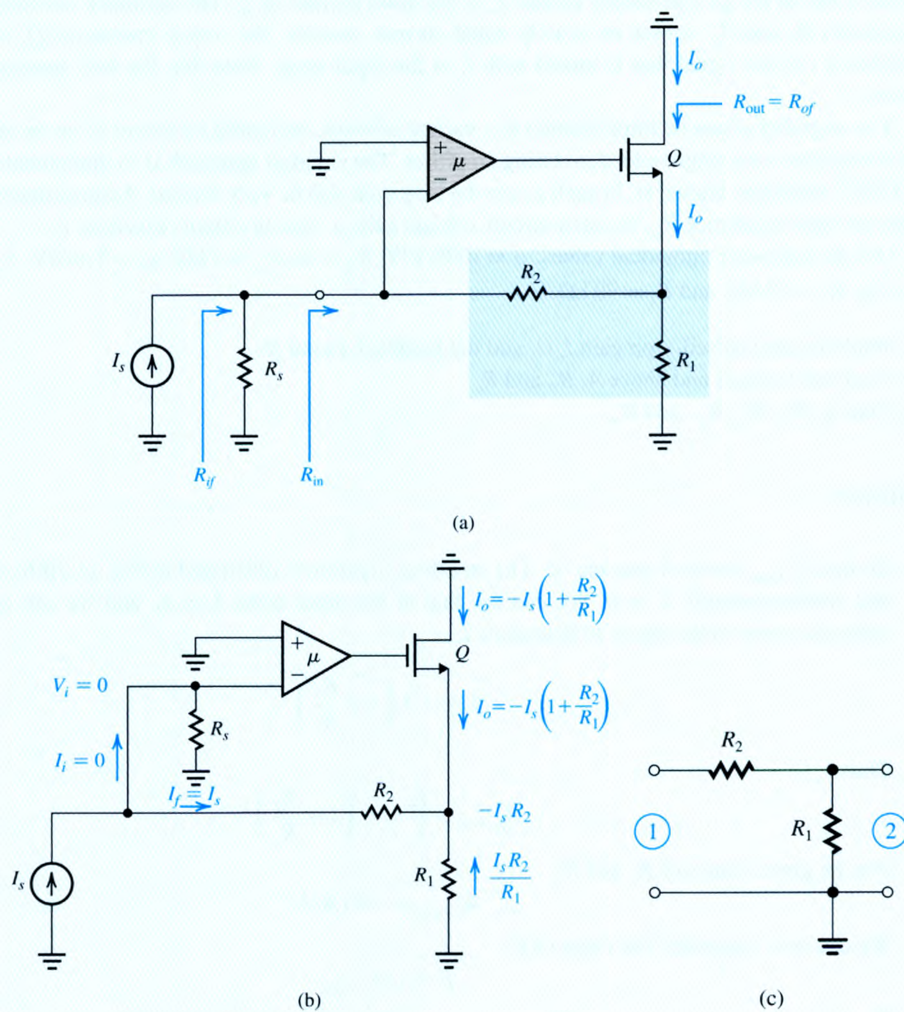


Figure 11.30 Circuits for Example 11.10.

with its equivalent circuit. Analysis of the A circuit is straightforward and proceeds as follows:

$$R_i = R_s \,\|\, R_{id} \,\|\, (R_1 + R_2)$$

$$V_i = I_i R_i$$

$$I_o = -\mu V_i \frac{1}{1/g_m + (R_1 \,\|\, R_2 \,\|\, r_{o2})} \frac{r_{o2}}{r_{o2} + (R_1 \,\|\, R_2)}$$

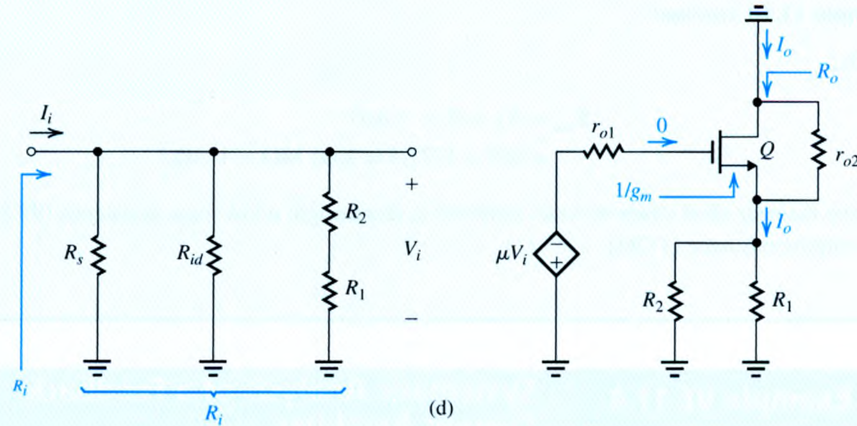


Figure 11.30 *continued*

Combining these three equations results in A:

$$A \equiv \frac{I_o}{I_i} = -\mu \frac{R_i}{1/g_m + (R_1 \| R_2 \| r_{o2})} \frac{r_{o2}}{r_{o2} + (R_1 \| R_2)}$$

Noting that R_o is the output resistance of Q, which has a resistance $(R_1 \| R_2)$ in its source lead, we can write

$$R_o = r_{o2} + (R_1 \| R_2) + (g_m r_{o2})(R_1 \| R_2)$$

Substituting the given numerical values gives us

$$A = -10{,}764 \text{ A/A}$$
$$R_i = 100 \text{ k}\Omega$$
$$R_o = 929 \text{ k}\Omega$$

(c)

$$A_f = \frac{A}{1 + A\beta} = \frac{-10{,}764}{1 + (-10{,}764) \times (-0.1)}$$
$$= -\frac{10{,}764}{1 + 1076.4} = -9.991 \text{ A/A}$$

which is very close to the ideal value of -10 A/A because of the large loop gain available; $A\beta = 1076.4$.

$$R_{if} = \frac{R_i}{1 + A\beta} = \frac{100 \text{ k}\Omega}{1077.4} = 92.8 \ \Omega$$

Since $R_s = \infty$,

$$R_{\text{in}} = R_{if} = 92.8 \ \Omega$$

Example 11.10 *continued*

Finally,

$$R_{out} = R_{of} = R_o(1 + A\beta)$$
$$= 929 \times 1077.4 = 1001 \text{ M}\Omega \simeq 1 \text{ G}\Omega$$

Observe the near ideal characteristics achieved in this design: a low input resistance (92.8 Ω) and a very high output resistance (1 GΩ).

Video Example VE 11.4 Systematic Analysis of a Feedback Current Amplifier

The feedback current amplifier in Fig. VE11.4 utilizes two identical NMOS transistors sized so that at $I_D = 0.1$ mA they operate at $V_{OV} = 0.2$ V. Both devices have $V_t = 0.5$ V and $V_A = 10$ V.

(a) If I_s has zero dc component, show that both Q_1 and Q_2 are operating at $I_D = 0.1$ mA. What is the dc voltage at the input?
(b) Find the ideal value of $A_f = I_o/I_s$, and the value of β.
(c) Find g_m and r_o for each of Q_1 and Q_2.
(d) Find the A circuit and the value of R_i, A, and R_o.
(e) Find $A\beta$ and A_f.
(f) Find R_{in} and R_{out}.

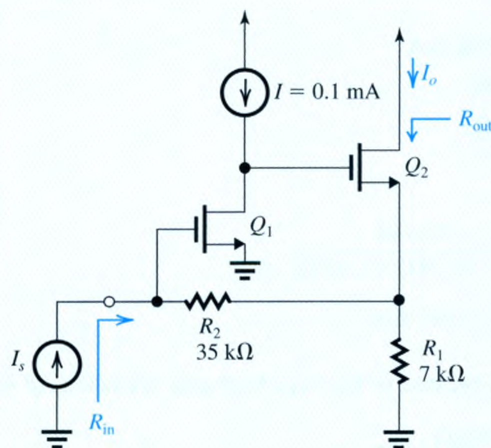


Figure VE11.4 Circuit for Video Example 11.4.


VE 11.4

Solution: Go to www.oup.com/he/sedra-smith8e to watch the authors solve this problem.

Related end-of-chapter problem: 11.74

EXERCISES

11.16 For the amplifier in Example 11.10, find the values of A_f, R_{in}, and R_{out} when the value of μ is 10 times lower, that is, when $\mu = 100$.
Ans. -9.91 A/A; 921 Ω; 101 MΩ

11.17 If in the circuit in Fig. 11.30(a), R_2 is short-circuited, find the ideal value of A_f. For the case $R_s = R_{id} = \infty$, give expressions for R_i, R_o, A, β, A_f, R_{in}, and R_{out}.
Ans. $A_f = -1$ A/A; $R_i = R_1$; $R_o = r_{o2}$, $A = -\mu g_m R_1$; $\beta = -1$; $A_f = -\mu g_m R_1/(1 + \mu g_m R_1)$; $R_{in} \simeq 1/\mu g_m$; $R_{out} \simeq \mu(g_m r_{o2})R_1$.

11.6 Summary of the Feedback-Analysis Method

Table 11.2 (found in the Summary Tables supplement online at www.oup.com/he/sedra-smith8e) summarizes the rules and relationships used in the analysis and design of the four types of feedback amplifier. In addition to the wealth of information found there, we offer the following important tips:

1. Always begin the analysis by determining the ideal value of the closed-loop gain A_f. This is done by assuming the open-loop gain $A = \infty$, which results in $V_i = 0$ and $I_i = 0$. Straightforward analysis is then performed to find $A_f\,|_{\text{ideal}} \equiv x_o/x_s$. The feedback factor is then found as $\beta = 1/A_f\,|_{\text{ideal}}$. Note also that with V_i and I_i equal to zero, the feedback circuit will be isolated and clearly evident.
2. Sketch the two-port feedback circuit and determine its loading effects R_{11} and R_{22}.
3. Sketch the A circuit and analyze it to determine A, R_i, and R_o.
4. Calculate $A\beta$, which must always be a positive dimensionless number, and the amount of feedback $(1 + A\beta)$.
5. Use the feedback formulas to determine A_f, R_{if}, and R_{of}. Check A_f versus the ideal value found in step 1; the difference should be approximately $(100/A\beta)\%$.
6. Extract R_s from R_{if} and R_L from R_{of} to determine R_{in} and R_{out}, respectively.
7. In using negative feedback to improve the properties of an amplifier under design, the first step is to select the feedback topology appropriate for the application. Then you can ascertain the required amount of negative feedback $(1 + A\beta)$ using the fact that it is this quantity that determines the magnitude of improvement in the various amplifier parameters. Also, you can determine the feedback factor β from the required closed-loop gain A_f; $\beta = 1/A_f\big|_{\text{ideal}}$.

11.7 The Stability Problem

Negative-feedback amplifiers suffer from a serious problem: the amplifier may become *unstable* and *oscillate*–that is, generate signals on its own, a clearly undesirable behavior!

In the remainder of this chapter, we study why and how this unstable behavior can happen, and most importantly what to do to ensure stable operation of the feedback amplifier.

In a feedback amplifier such as that represented by the general structure of Fig. 11.1, the open-loop gain A is generally a function of frequency, and it should therefore be more accurately called the **open-loop transfer function**, $A(s)$. Also, we have been assuming for the most part that the feedback network is resistive and hence that the feedback factor β is constant, but this need not be always the case. We shall therefore assume that in the general case the **feedback transfer function** is $\beta(s)$. It follows that the **closed-loop transfer function** $A_f(s)$ is given by

$$A_f(s) = \frac{A(s)}{1 + A(s)\beta(s)} \tag{11.29}$$

To focus attention on the points central to our discussion in this section, we shall assume that the amplifier is direct coupled with constant dc gain A_0 and with poles and zeros occurring in the high-frequency band. Also, for the time being let us assume that at low frequencies, $\beta(s)$ reduces to a constant value. Thus at low frequencies the loop gain $A(s)\beta(s)$ becomes a constant, which should be a positive number; otherwise the feedback would not be negative. The question then is: What happens at higher frequencies?

For physical frequencies $s = j\omega$, Eq. (11.29) becomes

$$A_f(j\omega) = \frac{A(j\omega)}{1 + A(j\omega)\beta(j\omega)} \tag{11.30}$$

Thus the loop gain $A(j\omega)\beta(j\omega)$ is a complex number that can be represented by its magnitude and phase,

$$A(j\omega)\beta(j\omega) = |A(j\omega)\beta(j\omega)|e^{j\phi(\omega)} \tag{11.31}$$

It is the manner in which the loop gain varies with frequency that determines the stability or instability of the feedback amplifier. To appreciate this fact, consider the frequency at which the phase angle $\phi(\omega)$ becomes 180°. At this frequency, ω_{180}, the loop gain $A(j\omega)\beta(j\omega)$ will be a real number with a negative sign. Thus at this frequency the feedback will become positive. If at $\omega = \omega_{180}$ the magnitude of the loop gain is less than unity, then from Eq. (11.30) we see that the closed-loop gain $A_f(j\omega)$ will be greater than the open-loop gain $A(j\omega)$, since the denominator of Eq. (11.30) will be smaller than unity. Nevertheless, the feedback amplifier will be stable.

On the other hand, if at the frequency ω_{180} the magnitude of the loop gain is equal to unity, it follows from Eq. (11.30) that $A_f(j\omega)$ will be infinite. This means that the amplifier will have an output for zero input; this is by definition an **oscillator**. To visualize how this feedback loop may oscillate, consider the general loop of Fig. 11.1 with the external input x_s set to zero. Any disturbance in the circuit, such as the closure of the power-supply switch, will generate a signal $x_i(t)$ at the input to the amplifier. Such a noise signal usually contains a wide range of frequencies, and we shall now concentrate on the component with frequency $\omega = \omega_{180}$, that is, the signal $X_i \sin(\omega_{180}t)$. This input signal will result in a feedback signal given by

$$X_f = A(j\omega_{180})\beta(j\omega_{180})X_i = -X_i$$

Since X_f is further multiplied by -1 in the summer block at the input, we see that the feedback causes the signal X_i at the amplifier input to be *sustained*. That is, from this point on, there will be sinusoidal signals at the amplifier input and output of frequency ω_{180}. Thus the amplifier is said to oscillate at the frequency ω_{180}.

The question now is: What happens if at ω_{180} the magnitude of the loop gain is greater than unity? We shall answer this question, not in general, but for the restricted yet very important class of circuits in which we are interested here. The answer, which is not obvious from Eq. (11.30), is that the circuit will oscillate, and the oscillations will grow in amplitude until some nonlinearity (which is always present in some form) reduces the magnitude of the loop gain to exactly unity, at which point sustained oscillations will be obtained. This mechanism for starting oscillations by using positive feedback with a loop gain greater than unity, and then using a nonlinearity to reduce the loop gain to unity at the desired amplitude, will be exploited in the design of sinusoidal oscillators in Chapter 15. Our objective here is just the opposite: Now that we know how oscillations could occur in a negative-feedback amplifier, we wish to find methods to prevent their occurrence.

Example 11.11

Consider a feedback amplifier for which the open-loop transfer function $A(s)$ is

$$A(s) = \left(\frac{10}{1+s/10^4}\right)^3$$

Let the feedback factor β be a constant independent of frequency. Find the frequency ω_{180} at which the phase shift of $A\beta$ is 180°. Then find the maximum value that β can have, β_{cr}, before the amplifier oscillates. Hence, find the *lowest* value of closed-loop gain for which the amplifier is stable.

Solution

The phase shift is

$$\phi = -3\tan^{-1}(\omega/10^4)$$

Thus, the frequency ω_{180} can be found from

$$-180° = -3\tan^{-1}(\omega_{180}/10^4)$$

$$\Rightarrow \omega_{180} = 10^4 \tan 60° = \sqrt{3} \times 10^4 \text{ rad/s}$$

At ω_{180},

$$|A| = \left(\frac{10}{\sqrt{1+3}}\right)^3 = 125$$

Thus, the loop gain will be

$$|A\beta| = 125\,\beta$$

For stable operation, we require

$$|A\beta| < 1$$

$$\Rightarrow 125\,\beta < 1$$

Example 11.11 *continued*

from which the maximum allowable β, denoted β_{cr}, is found as

$$\beta_{cr} = \frac{1}{125} = 0.008$$

It follows that for stable performance, the closed-loop gain cannot be lower than

$$A_f = \frac{10^3}{1 + 10^3\beta_{cr}} = \frac{10^3}{1 + 10^3 \times 0.008} = 111.1$$

11.8 Effect of Feedback on the Amplifier Poles

The frequency response and stability of an amplifier are determined directly by its poles. Therefore we shall investigate the effect of feedback on the poles of the amplifier.[7]

11.8.1 Stability and Pole Location

We begin by considering the relationship between stability and pole location. For an amplifier or any other system to be stable, its poles must lie in the left half of the s plane. A pair of complex-conjugate poles on the $j\omega$ axis gives rise to sustained sinusoidal oscillations. Poles in the right half of the s plane give rise to growing oscillations.

To verify the statement above, consider an amplifier with a pole pair at $s = \sigma_0 \pm j\omega_n$. If this amplifier is subjected to a disturbance, such as that caused by closure of the power-supply switch, its transient response will contain terms of the form

$$v(t) = e^{\sigma_0 t}\left[e^{+j\omega_n t} + e^{-j\omega_n t}\right] = 2e^{\sigma_0 t}\cos(\omega_n t) \tag{11.32}$$

This is a sinusoidal signal with an envelope $e^{\sigma_0 t}$. Now if the poles are in the left half of the s plane, then σ_0 will be negative and the oscillations will decay exponentially toward zero, as shown in Fig. 11.31(a), indicating that the system is stable. If, on the other hand, the poles are in the right half-plane, then σ_0 will be positive, and the oscillations will grow exponentially (until some nonlinearity limits their growth), as shown in Fig. 11.31(b). Finally, if the poles are on the $j\omega$ axis, then σ_0 will be zero and the oscillations will be sustained, as shown in Fig. 11.31(c).

Although the discussion above is in terms of complex-conjugate poles, it can be shown that the existence of any right-half-plane poles results in instability.

11.8.2 Poles of the Feedback Amplifier

From the closed-loop transfer function in Eq. (11.29), we see that the poles of the feedback amplifier are the zeros of $1 + A(s)\beta(s)$. That is, the feedback amplifier poles are obtained by

[7]For a brief review of poles and zeros and related concepts, refer to Appendix F.

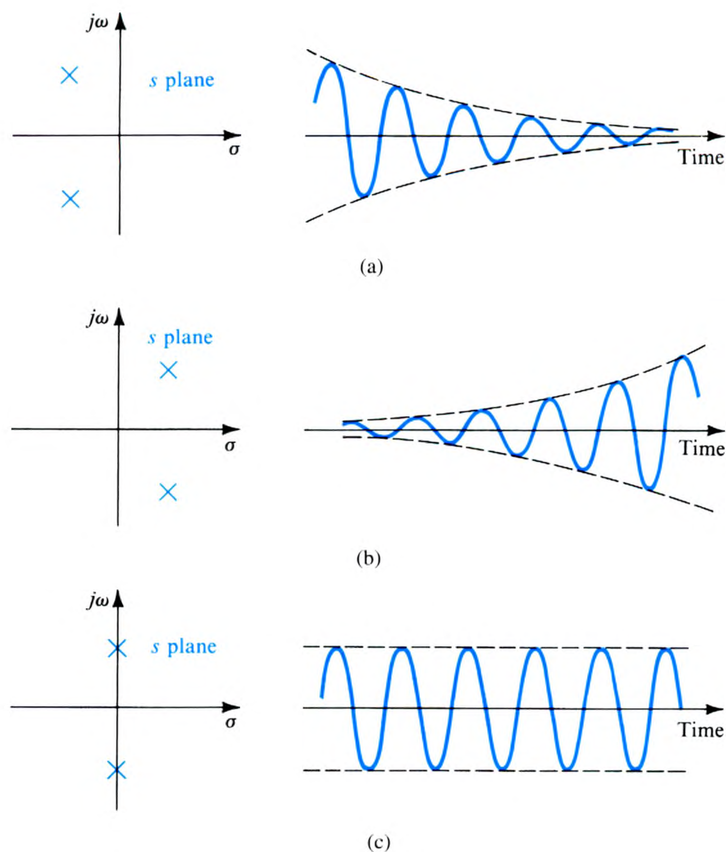


Figure 11.31 Relationship between pole location and transient response.

solving the equation

$$1 + A(s)\beta(s) = 0 \tag{11.33}$$

which is called the **characteristic equation** of the feedback loop. It should therefore be apparent that applying feedback to an amplifier changes its poles.

In the following, we shall consider how feedback affects the amplifier poles. For this purpose we shall assume that the open-loop amplifier has real poles and no finite zeros (i.e., all the zeros are at $s = \infty$). This will simplify the analysis and enable us to focus our attention on the fundamental concepts involved. We shall also assume that the feedback factor β is independent of frequency.

11.8.3 Amplifiers with a Single-Pole Response

Consider first the case of an amplifier whose open-loop transfer function is characterized by a single pole:

$$A(s) = \frac{A_0}{1 + s/\omega_P} \tag{11.34}$$

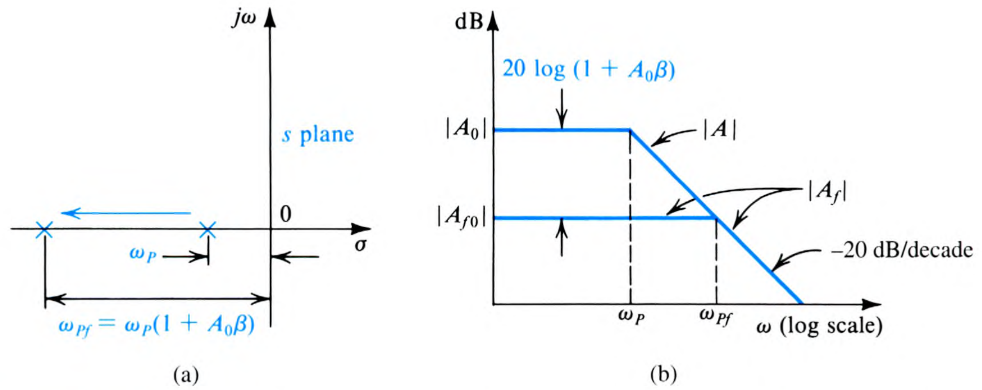


Figure 11.32 Effect of feedback on **(a)** the pole location and **(b)** the frequency response of an amplifier having a single-pole, open-loop response.

We can easily show that the closed-loop transfer function is given by

$$A_f(s) = \frac{A_0/(1+A_0\beta)}{1+s/\omega_P(1+A_0\beta)} \tag{11.35}$$

Thus the feedback moves the pole along the negative real axis to a frequency ω_{Pf},

$$\omega_{Pf} = \omega_P(1+A_0\beta) \tag{11.36}$$

This process is illustrated in Fig. 11.32(a). Figure 11.32(b) shows Bode plots for $|A|$ and $|A_f|$. Note that while at low frequencies the difference between the two plots is 20 log$(1 + A_0\beta)$, the two curves coincide at high frequencies. One can show that this indeed is the case by approximating Eq. (11.35) for frequencies $\omega \gg \omega_P(1 + A_0\beta)$:

$$A_f(s) \simeq \frac{A_0\omega_P}{s} \simeq A(s) \tag{11.37}$$

Physically speaking, at such high frequencies the loop gain is much smaller than unity and the feedback is ineffective.

Figure 11.32(b) clearly shows that applying negative feedback to an amplifier extends its bandwidth at the expense of a reduction in gain. Since the pole of the closed-loop amplifier never enters the right half of the s plane, the single-pole amplifier is stable for any value of β. Thus this amplifier is said to be **unconditionally stable**. This result, however, is hardly surprising, since the phase lag associated with a single-pole response can never be greater than 90°. Thus the loop gain never achieves the 180° phase shift required for the feedback to become positive.

EXERCISE

11.18 An op amp having a single-pole rolloff at 100 Hz and a low-frequency gain of 10^5 is operated in a feedback loop with $\beta = 0.01$. What is the factor by which feedback shifts the pole? To what frequency? If β is changed to a value that results in a nominal closed-loop gain of +1, to what frequency does the pole shift?
Ans. 1001; 100.1 kHz; 10 MHz

11.8.4 Amplifiers with a Two-Pole Response

Consider next an amplifier whose open-loop transfer function is characterized by two real-axis poles:

$$A(s) = \frac{A_0}{(1 + s/\omega_{P1})(1 + s/\omega_{P2})} \tag{11.38}$$

In this case, the closed-loop poles are obtained from $1 + A(s)\beta = 0$, which leads to

$$s^2 + s(\omega_{P1} + \omega_{P2}) + (1 + A_0\beta)\omega_{P1}\omega_{P2} = 0 \tag{11.39}$$

Thus the closed-loop poles are given by

$$s = -\frac{1}{2}(\omega_{P1} + \omega_{P2}) \pm \frac{1}{2}\sqrt{(\omega_{P1} + \omega_{P2})^2 - 4(1 + A_0\beta)\omega_{P1}\omega_{P2}} \tag{11.40}$$

From Eq. (11.40) we see that as the loop gain $A_0\beta$ is increased from zero, the poles are brought closer together. Then a value of loop gain is reached at which the second term in Eq. (11.40) becomes zero and the poles become coincident. If the loop gain is further increased, the poles become complex conjugate and move along a vertical line. Figure 11.33 shows the locus of the poles for increasing loop gain. This plot is called a **root-locus diagram**, where "root" refers to the fact that the poles are the roots of the characteristic equation.

From the root-locus diagram of Fig. 11.33 we see that this feedback amplifier also is unconditionally stable. Again, this result should come as no surprise; the maximum phase shift of $A(s)$ in this case is 180° (90° per pole), but this value is reached at $\omega = \infty$. Thus there is no finite frequency at which the phase shift reaches 180°.

Another observation to make on the root-locus diagram of Fig. 11.33 is that the open-loop amplifier might have a dominant pole, but this is not necessarily the case for the closed-loop amplifier. The response of the closed-loop amplifier can, of course, always be plotted once the poles have been found from Eq. (11.40). As is the case with second-order responses generally, the closed-loop response can show a peak (see Chapter 14). To be more specific, the characteristic equation of a second-order network can be written in the

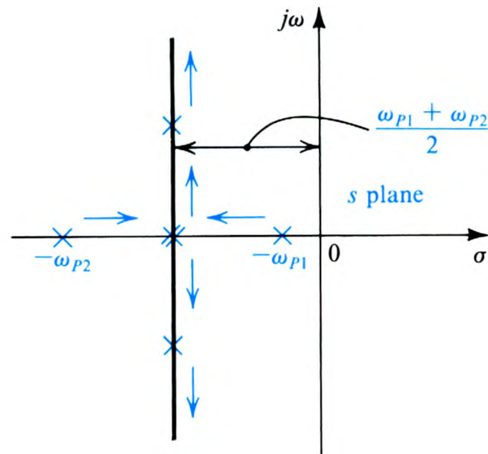


Figure 11.33 Root-locus diagram for a feedback amplifier whose open-loop transfer function has two real poles.

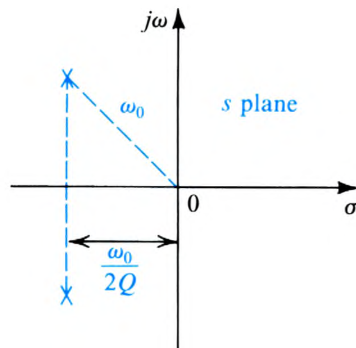


Figure 11.34 Definition of ω_0 and Q of a pair of complex-conjugate poles.

standard form

$$s^2 + s\frac{\omega_0}{Q} + \omega_0^2 = 0 \tag{11.41}$$

where ω_0 is called the **pole frequency** and Q is called **pole Q factor**. The poles are complex if Q is greater than 0.5. A geometric interpretation for ω_0 and Q of a pair of complex-conjugate poles is given in Fig. 11.34, from which we note that ω_0 is the radial distance of the poles from the origin and that Q indicates the distance of the poles from the $j\omega$ axis. Poles on the $j\omega$ axis have $Q = \infty$.

By comparing Eqs. (11.39) and (11.41), we obtain the Q factor for the poles of the feedback amplifier as

$$Q = \frac{\sqrt{(1 + A_0\beta)\omega_{P1}\omega_{P2}}}{\omega_{P1} + \omega_{P2}} \tag{11.42}$$

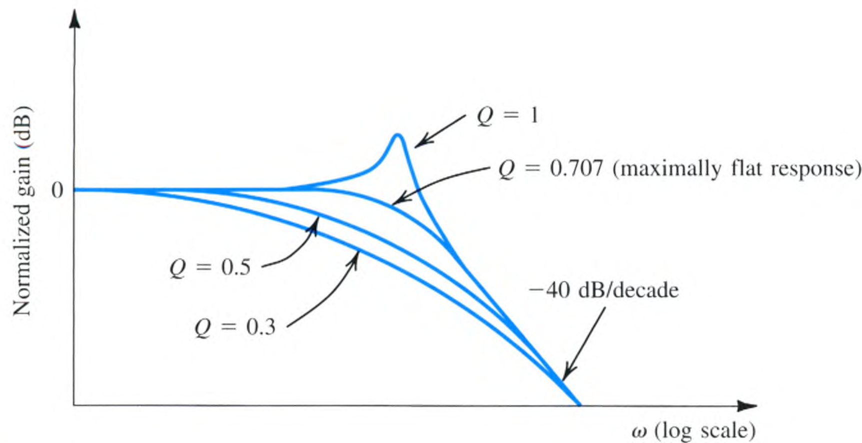


Figure 11.35 Normalized gain of a two-pole feedback amplifier for various values of Q. Note that Q is determined by the loop gain according to Eq. (11.42).

From the study of second-order network responses in Chapter 14, it will be seen that the response of the feedback amplifier under consideration shows no peaking for $Q \leq 0.707$. The boundary case corresponding to $Q = 0.707$ (poles at 45° angles) results in the **maximally flat** response. Figure 11.35 shows a number of possible responses obtained for various values of Q (or, correspondingly, various values of $A_0\beta$).

EXERCISE

11.19 An amplifier with a low-frequency gain of 100 and poles at 10^4 rad/s and 10^6 rad/s is incorporated in a negative-feedback loop with feedback factor β. For what value of β do the poles of the closed-loop amplifier coincide? What is the corresponding Q of the resulting second-order system? For what value of β is a maximally flat response achieved? What is the low-frequency closed-loop gain in the maximally flat case?

Ans. 0.245; 0.5; 0.5; 1.96 V/V

11.8.5 Amplifiers with Three or More Poles

Figure 11.36 shows the root-locus diagram for a feedback amplifier whose open-loop response is characterized by three poles. As indicated, increasing the loop gain from zero moves the highest-frequency pole outward while the two other poles are brought closer together. As $A_0\beta$ is increased further, the two poles become coincident and then become complex and conjugate. A value of $A_0\beta$ exists at which this pair of complex-conjugate poles enters the right half of the s plane, thus causing the amplifier to become unstable.

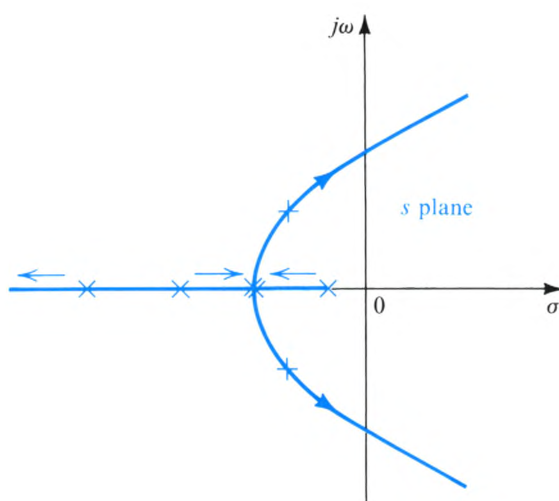


Figure 11.36 Root-locus diagram for an amplifier with three poles. The arrows indicate the pole movement as $A_0\beta$ is increased.

This result is not entirely unexpected, since an amplifier with three poles has a phase shift that reaches –270° as ω approaches ∞. Thus there exists a finite frequency, ω_{180}, at which the loop gain has 180° phase shift.

From the root-locus diagram of Fig. 11.36, we observe that one can always maintain amplifier stability by keeping the loop gain $A_0\beta$ smaller than the value corresponding to the poles entering the right half-plane. That is, for a given open-loop gain A_0, there exists a *maximum value* for β above which the feedback amplifier becomes unstable. Alternatively, we can state that there exists a *minimum value* for the closed-loop gain A_{f0} below which the amplifier becomes unstable. To obtain lower values of closed-loop gain, one needs therefore to alter the loop transfer function $A(s)\beta$. This is the process known as *frequency compensation*. We shall study the theory and techniques of frequency compensation in Section 11.10.

Before leaving this section, we point out that construction of the root-locus diagram for amplifiers having three or more poles as well as finite zeros is an involved process for which a systematic procedure exists. However, such a procedure will not be presented here, and the interested reader may consult Haykin (1970). Although the root-locus diagram provides the amplifier designer with considerable insight, other, simpler techniques based on Bode plots can be effectively employed, as we will explain in Section 11.9.

EXERCISE

11.20 Consider a feedback amplifier for which the open-loop transfer function $A(s)$ is given by

$$A(s) = \left(\frac{10}{1 + s/10^4}\right)^3$$

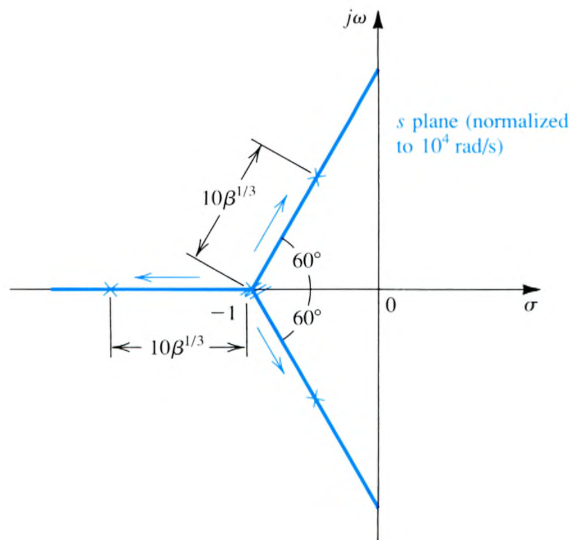


Figure E11.20

Let the feedback factor β be frequency independent. Find the closed-loop poles as functions of β, and show that the root locus is that of Fig. E11.20. Also find the value of β at which the amplifier becomes unstable. (*Note:* This is the same amplifier that was considered in Example 11.11.)

Ans. See Fig. E11.20; $\beta_{\text{critical}} = 0.008$

11.9 Stability Study Using Bode Plots

11.9.1 Gain and Phase Margins

From Section 11.7 we know that whether a feedback amplifier is or is not stable can be determined by examining its loop gain $A\beta$ as a function of frequency. One of the simplest and most effective means for doing this is through the use of a Bode plot for $A\beta$, such as the one shown in Fig. 11.37. (Note that because the phase approaches –360°, the circuit examined is a fourth-order one.) The feedback amplifier whose loop gain is plotted in Fig. 11.37 will be stable, since at the frequency of 180° phase shift, ω_{180}, the magnitude of the loop gain is less than unity (negative dB). The difference between the value of $|A\beta|$ at ω_{180} and unity, called the **gain margin**, is usually expressed in decibels. The gain margin represents the amount by which the loop gain can be increased while stability is maintained. Feedback amplifiers are usually designed to have sufficient gain margin to allow for the inevitable changes in loop gain with temperature, time, and so on.

Another way to investigate the stability and to express its degree is to examine the Bode plot at the frequency for which $|A\beta| = 1$, which is the point at which the magnitude plot crosses the 0-dB line. If at this frequency, ω_1, the phase angle is less (in magnitude) than 180°, then the amplifier is stable. This is the situation illustrated in Fig. 11.37. The difference between the phase angle at this frequency and 180° is termed the **phase margin**. On the other hand, if at the frequency of unity loop-gain magnitude, the phase lag is in excess of 180°, the amplifier will be unstable.

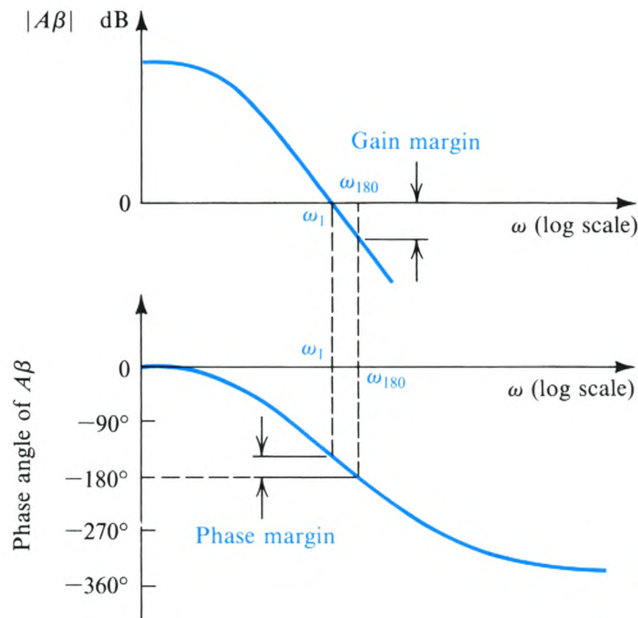


Figure 11.37 Bode plot for the loop gain $A\beta$ illustrating the definitions of the gain and phase margins.

EXERCISE

11.21 Consider an op amp having a single-pole, open-loop response with $A_0 = 10^5$ and $f_P = 10$ Hz. Let the op amp be ideal otherwise (infinite input impedance, zero output impedance, etc.). If this amplifier is connected in the noninverting configuration with a nominal low-frequency, closed-loop gain of 100, find the frequency at which $|A\beta| = 1$. Also, find the phase margin.
Ans. 10^4 Hz; 90°

11.9.2 Effect of Phase Margin on Closed-Loop Response

Feedback amplifiers are normally designed with a phase margin of at least 45°. The amount of phase margin has a profound effect on the shape of the closed-loop gain response. To derive this relationship, refer to Fig. 11.37. We see that at ω_1,

$$A(j\omega_1)\beta = 1 \times e^{-j\theta} \tag{11.43}$$

where

$$\theta = 180° - \text{phase margin} \tag{11.44}$$

At ω_1 the closed-loop gain is

$$A_f(j\omega_1) = \frac{A(j\omega_1)}{1 + A(j\omega_1)\beta} \tag{11.45}$$

Substituting from Eq. (11.43) gives

$$A_f(j\omega_1) = \frac{(1/\beta)e^{-j\theta}}{1+e^{-j\theta}} \tag{11.46}$$

Thus the magnitude of the gain at ω_1 is

$$|A_f(j\omega_1)| = \frac{1/\beta}{|1+e^{-j\theta}|} \tag{11.47}$$

Assuming that the feedback amplifier has a large low-frequency loop gain, that is, $A_0\beta \gg 1$, then

$$A_f(0) \simeq \frac{1}{\beta}$$

which enables us to write Equation (11.47) as

$$\frac{|A_f(j\omega_1)|}{A_f(0)} \simeq \frac{1}{|1+e^{-j\theta}|} \tag{11.48}$$

This relationship can be used to determine the relative peaking of the response for a given phase margin. For example, for a phase margin of 45°, $\theta = 135°$ and Eq. (11.48) provides

$$\frac{|A_f(j\omega_1)|}{A_f(0)} = 1.3$$

That is, the gain peaks by a factor of 1.3 above its low-frequency value. This peaking increases as the phase margin is reduced, eventually reaching ∞ when the phase margin is zero. Zero phase margin, of course, implies that the amplifier can sustain oscillations (poles on the $j\omega$ axis).

EXERCISE

11.22 Find the closed-loop gain at ω_1 relative to the low-frequency gain when the phase margin is 30°, 60°, and 90°.
Ans. 1.93; 1; 0.707

11.9.3 An Alternative Approach for Investigating Stability

Investigating stability by constructing Bode plots for the loop gain $A\beta$ can be a tedious and time-consuming process, especially if we have to investigate the stability of a given amplifier for a variety of feedback networks. An alternative approach, which is much simpler, is to construct a Bode plot for the open-loop gain $A(j\omega)$ only. Assuming for the time being that β is independent of frequency, we can plot 20 log$(1/\beta)$ as a horizontal straight line on the same plane used for 20 log$|A|$. The difference between the two curves will be

$$20\log|A(j\omega)| - 20\log\frac{1}{\beta} = 20\log|A\beta| \tag{11.49}$$

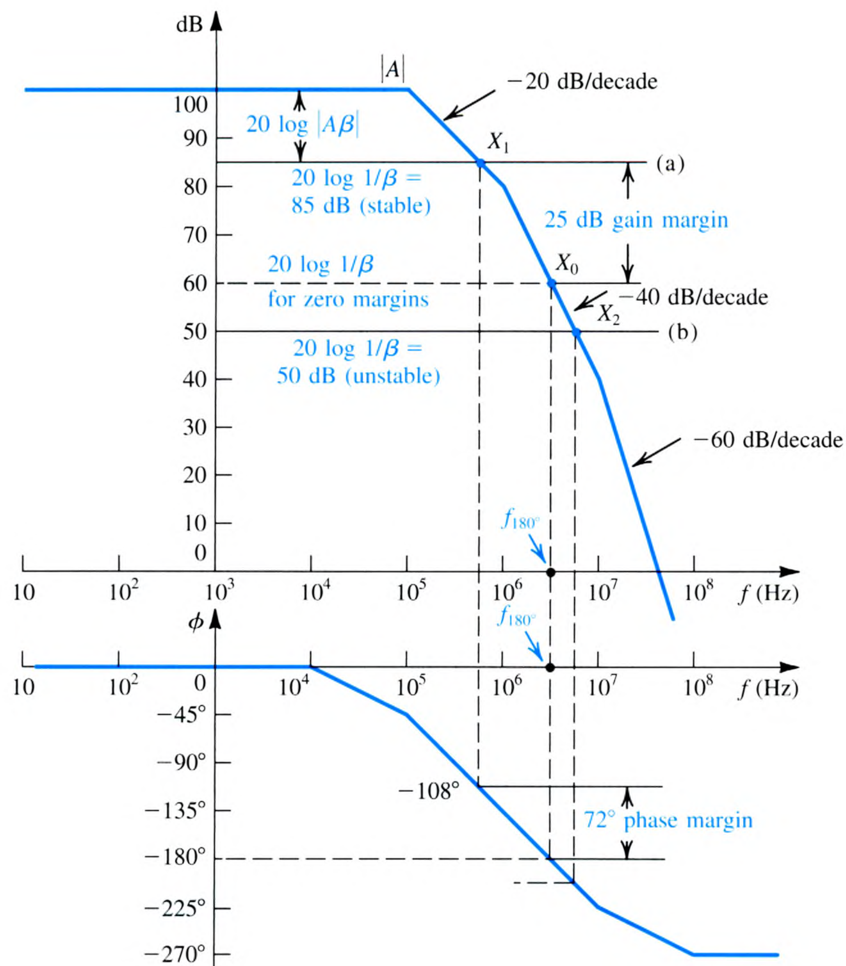


Figure 11.38 Stability analysis using Bode plot of A.

which is the loop gain (in dB). We may therefore study stability by examining the difference between the two plots. If we wish to evaluate stability for a different feedback factor, we simply draw another horizontal straight line at the level 20 log$(1/\beta)$. Note, specifically, that the point at which the horizontal straight line 20 log$(1/\beta)$ intersects the $|A|$ graph is the frequency ω_1 at which $|AB| = 1$. At this frequency, the magnitude of the amplifier phase must be less than 180° for the closed-loop amplifier to be stable.

To illustrate, consider an amplifier whose open-loop transfer function is characterized by three poles. For simplicity let the three poles be widely separated—say, at 0.1 MHz, 1 MHz, and 10 MHz, as shown in Fig. 11.38. Note that because the poles are widely separated, the phase is approximately –45° at the first pole frequency, –135° at the second, and –225° at the third. The frequency at which the phase of $A(j\omega)$ is –180° lies on the –40-dB/decade segment, as indicated in Fig. 11.38.

The open-loop gain of this amplifier can be expressed as

$$A = \frac{10^5}{(1 + jf/10^5)(1 + jf/10^6)(1 + jf/10^7)} \tag{11.50}$$

from which $|A|$ can be easily determined for any frequency f (in Hz), and the phase can be obtained as

$$\phi = -\left[\tan^{-1}(f/10^5) + \tan^{-1}(f/10^6) + \tan^{-1}(f/10^7)\right] \tag{11.51}$$

The magnitude and phase graphs shown in Fig. 11.38 are obtained using the method for constructing Bode plots (Appendix F). These graphs provide approximate values, with more exact values obtainable from Eqs. (11.50) and (11.51). For example, the frequency f_{180} at which the phase angle is 180° can be found from Fig. 11.38 to be approximately 3.2×10^6 Hz. Using this value as a starting point, a more exact value can be found by trial and error using Eq. (11.51). The result is $f_{180} = 3.34 \times 10^6$ Hz. At this frequency, Eq. (11.50) gives a gain magnitude of 58.2 dB, which is reasonably close to the approximate value of 60 dB given by Fig. 11.38.

To determine the maximum value of β, β_{cr}, at which the amplifier can be stable, albeit with zero margins, we locate f_{180} on the $|A|$ graph (point X_0) and draw a horizontal straight line. This will be the line at a height of approximately 60 dB. Thus, $20 \log(1/\beta_{cr}) = 60$ dB, which results in $\beta_{cr} = 10^{-3}$ and a corresponding closed-loop gain A_f of approximately 60 dB. Thus, the amplifier whose open-loop frequency response is shown in Fig. 11.38 *cannot* be used for closed-loop gains lower than 60 dB or 1000 V/V. In fact, even with a closed-loop gain of 60 dB, the feedback amplifier may oscillate since no margin is left to allow for possible change of gain.

Consider next the straight line labeled (a) in Fig. 11.38. This line represents a feedback factor for which $20 \log(1/\beta) = 85$ dB, which corresponds to $\beta = 5.623 \times 10^{-5}$ and a closed-loop gain of 83.6 dB. Since the loop gain is the difference between the $|A|$ curve and the $1/\beta$ line, the point of intersection X_1 corresponds to the frequency at which $|A\beta| = 1$. Using the graphs of Fig. 11.38, this frequency can be found to be approximately 5.6×10^5 Hz. (A more exact value of 4.936×10^5 can be obtained using the transfer function equations.) At this frequency the phase angle is approximately –108°. Thus the closed-loop amplifier, for which $20 \log(1/\beta) = 85$ dB, will be stable with a phase margin of 72°. The gain margin can be easily obtained from Fig. 11.38; it is 25 dB.

Next, suppose that we wish to use this amplifier to obtain a closed-loop gain of 50-dB nominal value. Since $A_0 = 100$ dB, we see that $A_0\beta \gg 1$ and $20 \log(A_0\beta) \simeq 50$ dB, resulting in $20 \log(1/\beta) \simeq 50$ dB. To verify that this closed-loop amplifier will not be stable, we draw line (b) in Fig. 11.38 with a height of 50 dB. This line intersects the open-loop gain curve at point X_2, where the corresponding phase is greater than 180°. Thus the closed-loop amplifier with 50-dB gain will be unstable.

Since the 180°-phase point always occurs on the –40-dB/decade segment of the Bode plot for $|A|$, a rule of thumb to guarantee stability is as follows: *The closed-loop amplifier will be stable if the* $20 \log(1/\beta)$ *line intersects the* $20 \log|A|$ *curve at a point on the* –20-dB/decade *segment.* Following this rule ensures that a phase margin of at least 45° is obtained.

The rule of thumb above can be generalized for the case in which β is a function of frequency. The general rule states that *at the intersection of* $20 \log\left[1/|\beta(j\omega)|\right]$ *and* $20 \log|A(j\omega)|$ *the difference of slopes* (called the **rate of closure**) *should not exceed* 20 dB/decade.

EXERCISE

11.23 For the amplifier whose open-loop-gain frequency response is shown in Fig. 11.38, find the value of β that results in a phase margin of 45°. What is the corresponding closed-loop gain?
Ans. 10^{-4}; 80 dB

11.10 Frequency Compensation

In the previous section we found that there is a minimum closed-loop gain for which the feedback amplifier is stable (60 dB for the amplifier in Fig. 11.38). The question naturally arises as to what to do if the desired closed-loop gain is lower than this minimum. The answer is to appropriately modify the open-loop frequency response $A(s)$. This process, known as *frequency compensation*, will be studied in this section.

11.10.1 Theory

The simplest method of frequency compensation consists of introducing a new pole in the function $A(s)$ at a frequency, f_D, sufficiently low that the modified open-loop gain, $A'(s)$, intersects the 20 $\log(1/|\beta|)$ curve with a slope difference of 20 dB/decade. For example, imagine that we are required to compensate the amplifier whose $A(s)$ is shown in Fig. 11.39 so that closed-loop amplifiers with β as high as 10^{-2} (i.e., closed-loop gains as low as approximately 40 dB) will be stable. First, we draw a horizontal straight line at the 40-dB level to represent 20 $\log(1/\beta)$, as shown in Fig. 11.39. We then locate point Y on this line at the frequency of the first pole, f_{P1}. From Y we draw a line with –20-dB/decade slope and determine the point at which this line intersects the dc gain line, point Y'. This latter point gives the frequency f_D of the new pole that has to be introduced in the open-loop transfer function.

The compensated open-loop response $A'(s)$ is indicated in Fig. 11.39. It has four poles: at f_D, f_{P1}, f_{P2}, and f_{P3}. Thus $|A'|$ begins to roll off with a slope of –20 dB/decade at f_D. At f_{P1} the slope changes to – 40 dB/decade, at f_{P2} it changes to – 60 dB/decade, and so on. Since the 20 $\log(1/\beta)$ line intersects the $20\log|A'|$ curve at point Y on the –20-dB/decade segment, the closed-loop amplifier with this β value (or lower values) will be stable.

A serious disadvantage of this compensation method is that at most frequencies the open-loop gain has been drastically reduced. This means that at most frequencies the amount of feedback available will be small. Since all the advantages of negative feedback are directly proportional to the amount of feedback, the performance of the compensated amplifier will be impaired.

A close look at Fig. 11.39 reveals that the gain $A'(s)$ is low because of the pole at f_{P1}. If we can somehow eliminate this pole, then—rather than locating point Y, drawing YY', and so on—we can start from point Z (at the frequency of the second pole) and draw the line ZZ'. This would result in the open-loop curve $A''(s)$, which shows much higher gain than $A'(s)$.

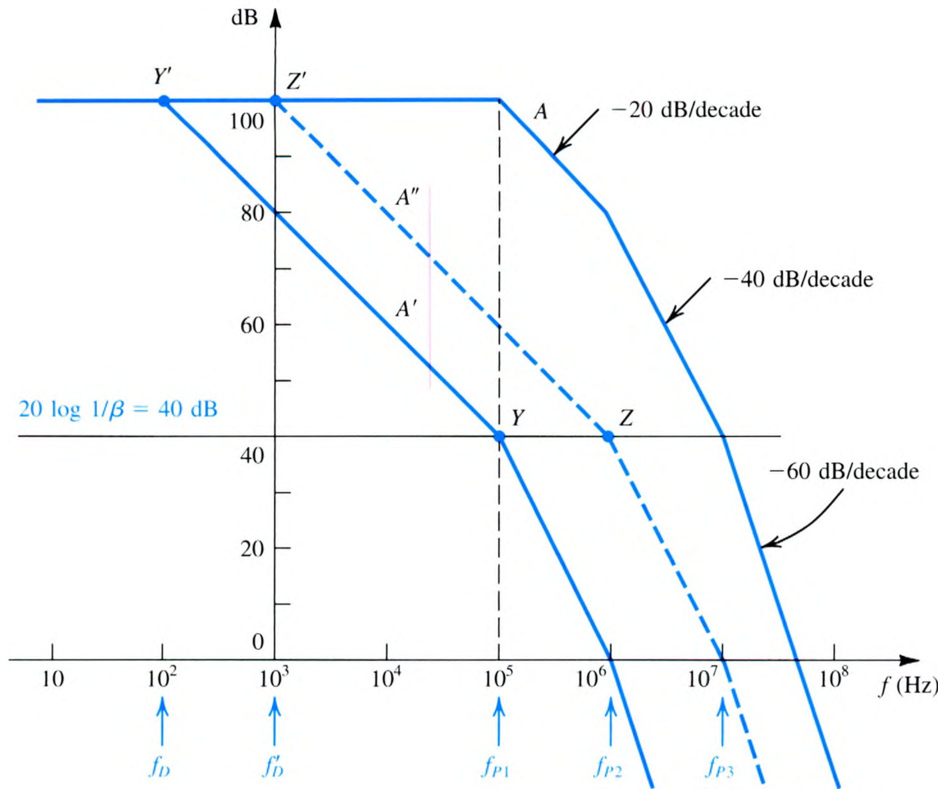


Figure 11.39 Frequency compensation for $\beta = 10^{-2}$. The response labeled A' is obtained by introducing an additional pole at f_D. The A'' response is obtained by moving the original low-frequency pole to f'_D.

Although it is not possible to eliminate the pole at f_{P1}, it is usually possible to shift that pole from $f = f_{P1}$ to $f = f'_D$. This makes the pole dominant and eliminates the need for introducing an additional lower-frequency pole, as will be explained next.

11.10.2 Implementation

We will now explain how to implement the frequency-compensation scheme discussed above. The amplifier circuit normally consists of a number of cascaded gain stages, with each stage responsible for one or more of the transfer function poles. Through manual and/or computer analysis of the circuit, we can identify the stage that introduces each of the important poles f_{P1}, f_{P2}, and so on. For now, let's assume that the first pole f_{P1} is introduced at the interface between the two cascaded differential stages shown in Fig. 11.40(a). In Fig. 11.40(b) we show a simple small-signal model of the circuit at this interface. Current source I_x represents the output-signal current of the Q_1–Q_2 stage. Resistance R_x and capacitance C_x represent the total resistance and capacitance between the two nodes B and B$'$. It follows that the pole f_{P1} is given by

$$f_{P1} = \frac{1}{2\pi C_x R_x} \tag{11.52}$$

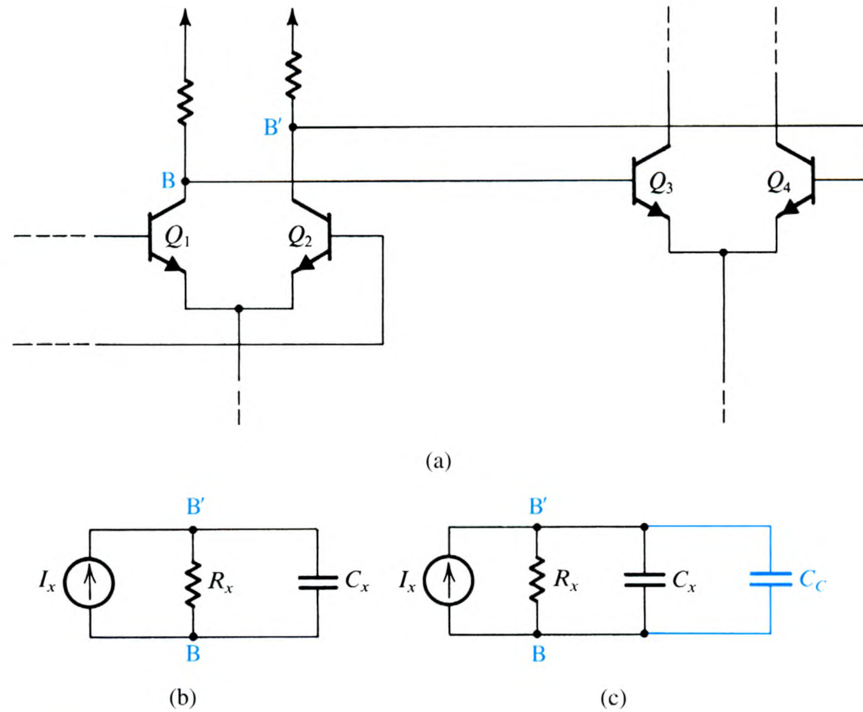


Figure 11.40 **(a)** Two cascaded gain stages of a multistage amplifier. **(b)** Equivalent circuit for the interface between the two stages in **(a)**. **(c)** Same circuit as in **(b)**, but with a compensating capacitor C_C added. Note that the analysis here applies equally well to MOS amplifiers.

Let us now connect the compensating capacitor C_C between nodes B and B′. This will result in the modified equivalent circuit shown in Fig. 11.40(c), from which we see that the pole introduced will no longer be at f_{P1}; rather, the pole can be at any desired lower frequency f'_D :

$$f'_D = \frac{1}{2\pi (C_x + C_C)R_x} \tag{11.53}$$

We can conclude that by selecting an appropriate value for C_C we can shift the pole frequency from f_{P1} to the value f'_D determined by point Z' in Fig. 11.39.

A disadvantage of this implementation method is that the required value of C_C is usually quite large. Thus if the amplifier to be compensated is an IC op amp, it will be difficult, and probably impossible, to include this compensating capacitor on the IC chip. (As pointed out in Chapter 8 and in Appendix A, the maximum practical size of a monolithic capacitor is about 100 pF.) An elegant solution to this problem is to connect the compensating capacitor in the feedback path of an inverting amplifier stage. Because of the Miller effect (Section 10.2), the compensating capacitance will be multiplied by the stage gain, resulting in a much larger effective capacitance. Furthermore, as explained later, another unexpected benefit accrues.

11.10.3 Miller Compensation and Pole Splitting

Figure 11.41(a) shows one gain stage in a multistage amplifier. For simplicity, the stage is shown as a common-emitter amplifier, but in practice it can be a more elaborate circuit. In the feedback path of this common-emitter stage we have placed a compensating capacitor C_f.

Figure 11.41(b) shows a simplified equivalent circuit of the gain stage of Fig. 11.41(a). Here R_1 and C_1 represent the total resistance and total capacitance between node B and ground. Similarly, R_2 and C_2 represent the total resistance and total capacitance between node C and ground. Furthermore, it is assumed that C_1 includes the Miller component due to capacitance C_μ, and C_2 includes the input capacitance of the succeeding amplifier stage. Finally, I_i represents the output signal current of the preceding stage.

In the absence of the compensating capacitor C_f, we can see from Fig. 11.41(b) that there are two poles—one at the input and one at the output. Let us assume that these two poles are f_{P1} and f_{P2} of Fig. 11.39; thus,

$$f_{P1} = \frac{1}{2\pi C_1 R_1} \qquad f_{P2} = \frac{1}{2\pi C_2 R_2} \tag{11.54}$$

With C_f present, analysis of the circuit yields the transfer function

$$\frac{V_o}{I_i} = \frac{(sC_f - g_m)R_1 R_2}{1 + s\left[C_1 R_1 + C_2 R_2 + C_f(g_m R_1 R_2 + R_1 + R_2)\right] + s^2\left[C_1 C_2 + C_f(C_1 + C_2)\right]R_1 R_2} \tag{11.55}$$

The zero is usually at a much higher frequency than the dominant pole, and we shall neglect its effect. The denominator polynomial $D(s)$ can be written in the form

$$D(s) = \left(1 + \frac{s}{\omega'_{P1}}\right)\left(1 + \frac{s}{\omega'_{P2}}\right) = 1 + s\left(\frac{1}{\omega'_{P1}} + \frac{1}{\omega'_{P2}}\right) + \frac{s^2}{\omega'_{P1}\omega'_{P2}} \tag{11.56}$$

where ω'_{P1} and ω'_{P2} are the new frequencies of the two poles. Normally one of the poles will be dominant; $\omega'_{P1} \ll \omega'_{P2}$. Thus,

$$D(s) \simeq 1 + \frac{s}{\omega'_{P1}} + \frac{s^2}{\omega'_{P1}\omega'_{P2}} \tag{11.57}$$

Equating the coefficients of s in the denominator of Eq. (11.55) and in Eq. (11.57) results in

$$\omega'_{P1} = \frac{1}{C_1 R_1 + C_2 R_2 + C_f(g_m R_1 R_2 + R_1 + R_2)}$$

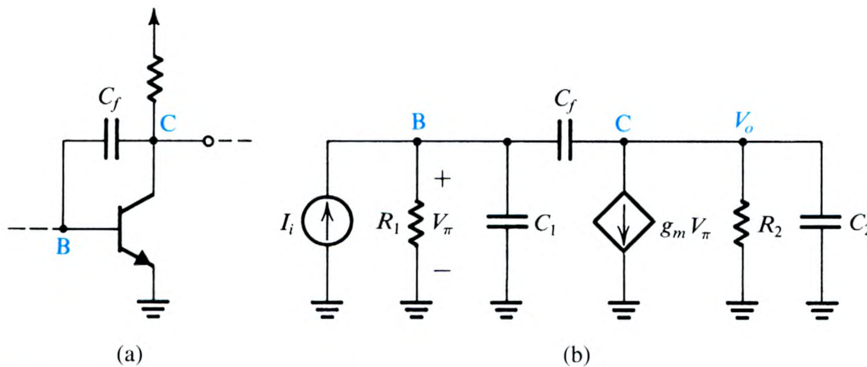


Figure 11.41 **(a)** A gain stage in a multistage amplifier with a compensating capacitor connected in the feedback path, and **(b)** equivalent circuit. Note that although a BJT is shown, the analysis applies equally well to the MOSFET case.

which can be approximated by

$$\omega'_{P1} \simeq \frac{1}{g_m R_2 C_f R_1} \tag{11.58}$$

To obtain ω'_{P2} we equate the coefficients of s^2 in the denominator of Eq. (11.55) and in Eq. (11.57) and use Eq. (11.58):

$$\omega'_{P2} \simeq \frac{g_m C_f}{C_1 C_2 + C_f(C_1 + C_2)} \tag{11.59}$$

From Eqs. (11.58) and (11.59), we see that as C_f is increased, ω'_{P1} is reduced and ω'_{P2} is increased. This action is referred to as **pole splitting**. Note that the increase in ω'_{P2} is highly beneficial; it allows us to move point Z (see Fig. 11.39) further to the right, thus resulting in higher compensated open-loop gain. Finally, note from Eq. (11.58) that C_f is multiplied by the Miller effect factor $g_m R_2$, thus resulting in a much larger effective capacitance, $g_m R_2 C_f$. In other words, the required value of C_f will be much smaller than that of C_C in Fig. 11.40.

Example 11.12

Consider an op amp whose open-loop transfer function is identical to that shown in Fig. 11.38. We wish to compensate this op amp so that the closed-loop amplifier with resistive feedback is stable for any gain (i.e., for β up to unity). Assume that the op-amp circuit includes a stage such as that of Fig. 11.41 with $C_1 = 100$ pF, $C_2 = 5$ pF, and $g_m = 40$ mA/V, that the pole at f_{P1} is caused by the input circuit of that stage, and that the pole at f_{P2} is introduced by the output circuit. Find the value of the compensating capacitor for two cases: if it is connected between the input node B and ground or if it is connected in the feedback path of the transistor.

Solution

First we determine R_1 and R_2 from

$$f_{P1} = 0.1 \text{ MHz} = \frac{1}{2\pi C_1 R_1}$$

Thus,

$$R_1 = \frac{10^5}{2\pi} \, \Omega$$

$$f_{P2} = 1 \text{ MHz} = \frac{1}{2\pi C_2 R_2}$$

Thus,

$$R_2 = \frac{10^5}{\pi} \, \Omega$$

If a compensating capacitor C_C is connected across the input terminals of the transistor stage, then the frequency of the first pole changes from f_{P1} to f'_D:

$$f'_D = \frac{1}{2\pi (C_1 + C_C) R_1}$$

The second pole remains unchanged at 1 MHz. The required value for f'_D is determined by drawing a –20-dB/decade line from the 1-MHz frequency point on the $20 \log(1/\beta) = 20 \log 1 = 0$ dB line. This line will intersect the 100-dB dc gain line at 10 Hz. Thus,

$$f'_D = 10\,\text{Hz} = \frac{1}{2\pi (C_1 + C_C) R_1}$$

which results in $C_C \simeq 1\ \mu\text{F}$, which is quite large and certainly cannot be included on the IC chip.

Next, if a compensating capacitor C_f is connected in the feedback path of the transistor, then both poles change location to the values given by Eqs. (11.58) and (11.59):

$$f'_{P1} \simeq \frac{1}{2\pi g_m R_2 C_f R_1} \qquad f'_{P2} \simeq \frac{g_m C_f}{2\pi [C_1 C_2 + C_f (C_1 + C_2)]} \tag{11.60}$$

To determine where we should locate the first pole, we need to know the value of f'_{P2}. As an approximation, let us assume that $C_f \gg C_2$, which enables us to obtain

$$f'_{P2} \simeq \frac{g_m}{2\pi (C_1 + C_2)} = 60.6\,\text{MHz}$$

Thus it appears that this pole will move to a frequency higher than f_{P3} (which is 10 MHz). Let us therefore assume that the second pole of the compensated amplifier will be at f_{P3}. This requires that the modified first pole be located at

$$f'_{P1} = \frac{f_{P3}}{A_0} = \frac{10^7\,\text{Hz}}{10^5} = 100\,\text{Hz}$$

Thus,

$$f'_{P1} = 100\,\text{Hz} = \frac{1}{2\pi g_m R_2 C_f R_1}$$

which results in $C_f = 78.5$ pF. Although this value is indeed much greater than C_2, we can determine the location of the pole f'_{P2} from Eq. (11.60), which yields $f'_{P2} = 57.2$ MHz, confirming that this pole has indeed been moved past f_{P3}.

We conclude that using Miller compensation not only results in a much smaller compensating capacitor but, owing to pole splitting, also enables us to place the dominant pole a decade higher in frequency. This results in a wider bandwidth for the compensated op amp.

EXERCISES

11.24 A multipole amplifier having a first pole at 1 MHz and a dc open-loop gain of 100 dB is to be compensated for closed-loop gains as low as 20 dB by the introduction of a new dominant pole. At what frequency must the new pole be placed?
Ans. 100 Hz

11.25 For the amplifier described in Exercise 11.24, rather than introducing a new dominant pole, we can use additional capacitance at the circuit node at which the first pole is formed to reduce the frequency of the first pole. If the frequency of the second pole is 10 MHz and if it remains unchanged while additional capacitance is introduced as mentioned, find the frequency to which the first pole must be lowered so that the resulting amplifier is stable for closed-loop gains as low as 20 dB. By what factor must the capacitance at the controlling node be increased?
Ans. 1000 Hz; 1000

Summary

- Negative feedback is employed to make the amplifier gain less sensitive to component variations; to control input and output resistances; to extend bandwidth; to reduce nonlinear distortion; and to enhance signal-to-interference ratio.
- The advantages above are obtained at the expense of a reduction in gain and at the risk of the amplifier becoming unstable (that is, oscillating). The latter problem is solved by careful design.
- The structure of an ideal negative-feedback amplifier is shown in Fig. 11.1. Table 11.1 (in the online Summary Tables supplement) summarizes the parameters and relationships governing the operation of the ideal structure.
- For each of the four basic types of amplifier, there is an appropriate feedback topology. The four topologies, together with their analysis procedure and their effects on input and output impedances, are summarized in Table 11.2 in the online Summary Tables supplement.
- The key feedback parameters are the loop gain ($A\beta$), which for negative feedback must be a positive dimensionless number, and the amount of feedback ($1 + A\beta$). The latter directly determines gain reduction, gain desensitivity, bandwidth extension, and changes in R_i and R_o.
- The loop gain $A\beta$ can be determined by breaking the feedback loop, as illustrated in Figs. 11.2 and 11.11. The value of $A\beta$ can be used together with the feedback factor β to determine A and hence A_f. This method, though simple, is incomplete as it does not enable the determination of the input and output resistances. For these, we utilize the systematic method for feedback analysis (refer to Table 11.2).
- A series connection at the input or output of a feedback amplifier results in increasing the input or output resistance, respectively, by $(1 + A\beta)$. A parallel or shunt connection reduces these resistances.
- Always begin the analysis by determining the ideal value of the closed-loop gain A_f. This is done by assuming the open-loop gain $A = \infty$ which results in $V_i = 0$ and $I_i = 0$. Straightforward analysis is then performed to find $A_f|_{\text{ideal}} = x_o/x_s$. The feedback factor is then found as $\beta = 1/A_f|_{\text{ideal}}$. Note also that with V_i and I_i equal to zero, the feedback circuit will be clearly evident.
- The ideal or upper-bound value of the closed-loop gain A_f is $1/\beta$ and is approached when $A\beta \gg 1$ and obtained for $A = \infty$.
- Since A and β are in general frequency dependent, the poles of the feedback amplifier are obtained by solving the characteristic equation $1 + A(s)\beta(s) = 0$.
- For the feedback amplifier to be stable, its poles must all be in the left half of the s plane.

- Stability is guaranteed if at the frequency for which the phase angle of $A\beta$ is 180° (i.e., ω_{180}), $|A\beta|$ is less than unity; the amount by which it is less than unity, expressed in decibels, is the gain margin. Alternatively, the amplifier is stable if, at the frequency at which $|A\beta| = 1$, the phase angle is less than 180°; the difference is the phase margin.
- The stability of a feedback amplifier can be analyzed by constructing a Bode plot for $|A|$ and superimposing on it a plot for $20\log 1/|\beta|$. Stability is guaranteed if the two plots intersect with a difference in slope no greater than 20 dB/decade.
- To make a given amplifier stable for a given feedback factor β, the open-loop frequency response is suitably modified by a process known as frequency compensation.
- A popular method for frequency compensation involves connecting a feedback capacitor across an inverting stage in the amplifier. This causes the pole formed at the input of the amplifier stage to shift to a lower frequency and thus become dominant, while the pole formed at the output of the amplifier stage is moved to a very high frequency and thus becomes unimportant. This process is known as pole splitting.

PROBLEMS

Computer Simulation Problems

SIM Problems identified by this icon are intended to demonstrate the value of using SPICE simulation to verify hand analysis and design, and to investigate important issues such as allowable signal swing and amplifier nonlinear distortion. Instructions to assist in setting up simulations for all the indicated problems can be found in the corresponding files on the website. Note that if a particular parameter value is not specified in the problem statement, you are to make a reasonable assumption.

Section 11.1: The General Feedback Structure

11.1 A negative-feedback amplifier has a closed-loop gain $A_f = 100$ and an open-loop gain $A = 10^4$. Find the values of the amount of feedback, the loop gain, and the feedback factor β. If a manufacturing error results in a reduction in A of 10%, what closed-loop gain results? What is the percentage change in A_f corresponding to this 10% reduction in A? Repeat for the case A changes to 10^3.

D 11.2 Design a feedback amplifier with an ideal closed-loop gain of 100. What is the required value of β? If the maximum deviation of the closed-loop gain from the ideal value is to be limited to 1%, what are the minimum required values of the loop gain and the open-loop gain?

11.3 Consider the op-amp circuit shown in Fig. P11.3, where the op amp has infinite input resistance and zero output resistance but finite open-loop gain A.

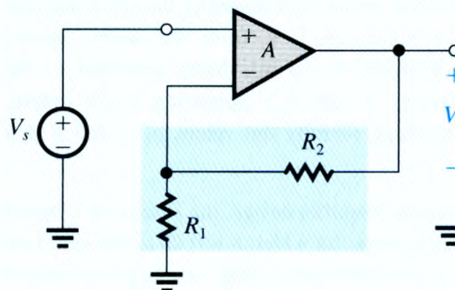


Figure P11.3

(a) By setting $A = \infty$, find expresssions for $A_f|_{\text{ideal}}$ and β.
(b) If $R_1 = 10\text{ k}\Omega$, find R_2 that results in $A_f = 10$ V/V for the following three cases: (i) $A = 1000$ V/V; (ii) $A = 100$ V/V; (iii) $A = 20$ V/V.

---

problems with blue numbers are considered essential; * = difficult problem; ** = more difficult; *** = very challenging
SIM = simulation; **D** = design problem; ● = see related video example

(c) For each of the three cases in (b), find the percentage change in A_f that results when A decreases by 10%. Comment on the results.

11.4 The noninverting buffer op-amp configuration shown in Fig. P11.4 provides a direct implementation of the feedback loop of Fig. 11.1. Assuming that the op amp has infinite input resistance and zero output resistance, what is β? If A = 1000, what is the closed-loop voltage gain? What is the amount of feedback (in dB)? For $V_s = 1$ V, find V_o and V_i. If A decreases by 10%, what is the corresponding percentage decrease in A_f?

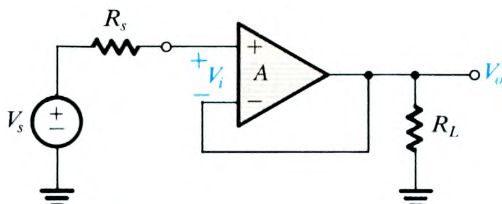


Figure P11.4

D 11.5 Design a feedback amplifier to have an ideal closed-loop gain of 10 and to have an amount of feedback of at least 40 dB. What is the required value of β? What is the minimum required value of the open-loop gain? What is the corresponding realized value of closed-loop gain?

11.6 In a particular circuit represented by the block diagram of Fig. 11.1, a signal of 1 V from the source results in a difference signal of 10 mV being provided to the amplifying element A, and 4 V appearing at the output. For this arrangement, identify the values of A and β that apply.

11.7 In a particular amplifier design, the β network consists of a linear potentiometer for which β is 0.00 at one end, 1.00 at the other end, and 0.50 in the middle. As the potentiometer is adjusted, find the three values of closed-loop gain that result when the amplifier open-loop gain is (a) 1 V/V, (b) 10 V/V, (c) 100 V/V, (d) 1000 V/V, and (e) 10,000 V/V. Provide your results in a table in which there is a row for each value of A and a column for each value of β.

11.8 A newly constructed feedback amplifier undergoes a performance test with the following results: With the feedback connection removed, a source signal of 2 mV is required to provide a 3-V output; with the feedback connected, a 3-V output requires a 100-mV source signal. For this amplifier, identify values for A, A_f, the amount of feedback, $A\beta$, and β.

D 11.9 An amplifier has an open-loop gain with a nominal value of 1000 but can vary from unit to unit by as much as ±50% of nominal. It is required to apply negative feedback to this amplifier so that the variability of the closed-loop gain of the resulting feedback amplifier is limited to ±1%. What is the largest possible nominal value of closed-loop gain that can be achieved? Now if three of these feedback amplifiers are placed in cascade, what is the nominal value of the gain of the resulting cascade amplifier? What is the expected variability of this gain?

Section 11.2: Some Properties of Negative Feedback

11.10 For the negative-feedback loop of Fig. 11.1, find the loop gain $A\beta$ for which the sensitivity of closed-loop gain to open-loop gain [i.e., $(dA_f/A_f)/(dA/A)$] is –40 dB. For what value of $A\beta$ does the sensitivity become 0.1?

D11.11 A designer is considering two possible designs of a feedback amplifier. The goal is $A_f = 10$ V/V. One design employs an amplifier for which $A = 1000$ V/V and the other uses $A = 500$ V/V. Find the desensitivity factor and β in both cases. If the $A = 1000$ amplifier units have a gain uncertainty of ±10%, what is the gain uncertainty for the closed-loop amplifiers utilizing this amplifier type? If the same result is to be achieved with the $A = 500$ amplifier, what is the maximum allowable uncertainty in its gain?

D11.12 A designer is required to achieve a closed-loop gain of 20 ± 0.1% V/V using a basic amplifier whose gain variation is ±10%. What nominal value of A and β (assumed constant) are required?

D11.13 A circuit designer requires a gain of 20 ± 1% V/V using an amplifier whose gain varies by a factor of 10 over temperature and time. What is the lowest open-loop gain required? The value of β? (*Hint*: Since the change in the open-loop gain is very large, do *not* use differential analysis.)

D*11.14 A power amplifier employs an output stage whose gain varies from 2 to 12 for various reasons. What is the gain of an ideal (nonvarying) amplifier connected to drive it

so that an overall gain with feedback of $100 \pm 5\%$ V/V can be achieved? What is the value of β to be used? What are the requirements if A_f must be held within $\pm 0.5\%$? For each of these situations, what preamplifier gain and feedback factor β are required if A_f is to be 10 V/V (with the two possible tolerances)? (*Hint*: Since the change in the open-loop gain is very large, do *not* use differential analysis.)

D*11.15 Design an amplifier with a gain of 100 that is accurate to within $\pm 1\%$. You have available amplifier stages with a gain of 1000 that is accurate to within $\pm 20\%$. Provide a design that uses a number of these gain stages in cascade, with each stage employing negative feedback of an appropriate amount. Obviously, your design should use the lowest possible number of stages while meeting specification. What is the actual gain range realized in your design?

D *11.16 Design an amplifier to have a nominal closed-loop gain of 10 V/V using a battery-operated amplifier whose gain reduces to half its normal full-battery value over the life of the battery. If only 1% drop in closed-loop gain is desired, what nominal open-loop amplifier gain must be used in the design? (Note that since the change in A is large, it is inaccurate to use differentials.) What value of β should be chosen? If component-value variation in the β network may produce as much as a $\pm 1\%$ variation in β, to what value must A be raised to ensure the required minimum gain?

D 11.17 Design a feedback amplifier that has a closed-loop gain of 100 V/V and is relatively insensitive to change in basic-amplifier gain. In particular, it should provide a reduction in A_f to 99 V/V for a reduction in A to one-tenth its nominal value. What is the required loop gain? What nominal value of A is required? What value of β should be used? What would the closed-loop gain become if A were increased tenfold? If A were made infinite?

11.18 An operational amplifier has a dc open-loop gain of 100 dB and a uniform −20 dB/decade gain rolloff with a 3-dB frequency of 100 Hz. Find the amount of feedback and the feedback factor required to obtain the following dc closed-loop gains. In each case specify the 3-dB frequency of the feedback amplifier and the gain–bandwidth product.

(a) 60 dB; (b) 40 dB; (c) 20 dB; and (d) 0 dB.

D 11.19 Design a feedback amplifier with a dc gain of 10 V/V and a 3-dB bandwidth of at least 10 MHz. The available open-loop amplifier has a dc gain of 1000 V/V and −20 dB/decade gain rolloff with frequency. Specify the minimum required f_{3dB} of the open-loop amplifier. What is the feedback factor needed?

11.20 Consider an amplifier having a midband gain A_M and a low-frequency response characterized by a pole at $s = -\omega_L$ and a zero at $s = 0$. Let the amplifier be connected in a negative-feedback loop with a feedback factor β. Find an expression for the midband gain and the lower 3-dB frequency of the closed-loop amplifier. By what factor have both changed?

D *11.21 Design an amplifier with a low-frequency gain of 1000 and a 3-dB frequency of 1 MHz. You have available gain stages with a gain of 1000 but with a dominant high-frequency pole at 20 kHz. Provide a design that employs a number of such stages in cascade, each with negative feedback of an appropriate amount. Use identical stages.

Hint: The 3-dB frequency of a cascade of N identical gain stages, each with a 3-dB frequency $f_{3dB}|_{stage}$ is given by

$$f_{3dB}|_{cascade} = f_{3dB}|_{stage}\sqrt{2^{1/N} - 1}$$

D 11.22 A particular amplifier has a nonlinear transfer characteristic that can be approximated as follows:

(a) For small input signals, $|v_I| \le 5$ mV, $v_O/v_I = 10^3$.
(b) For intermediate input signals, 5 mV $\le |v_I| \le 55$ mV, $\Delta v_O/\Delta v_I = 10^2$.
(c) For large input signals, $|v_I| \ge 55$ mV, the output saturates.

If the amplifier is connected in a negative-feedback loop, find the feedback factor β that reduces the factor-of-10 change in gain (occurring at $|v_I| = 5$ mV) to only a 10% change. What is the transfer characteristic v_O versus v_S of the amplifier with feedback?

***11.23** The complementary BJT follower shown in Fig. P11.23(a) has the approximate transfer characteristic shown in Fig. P11.23(b). Observe that for $-0.7\text{ V} \le v_I \le +0.7$ V, the output is zero. This "dead band" leads to crossover distortion (see Section 12.3). Consider this follower to be driven by the output of a differential amplifier of gain 100 whose positive-input terminal is connected to the input signal source v_S and whose negative-input terminal is connected to the emitters of the follower. Sketch the transfer characteristic v_O versus v_S of the resulting feedback

amplifier. What are the limits of the dead band, and what are the gains outside the dead band?

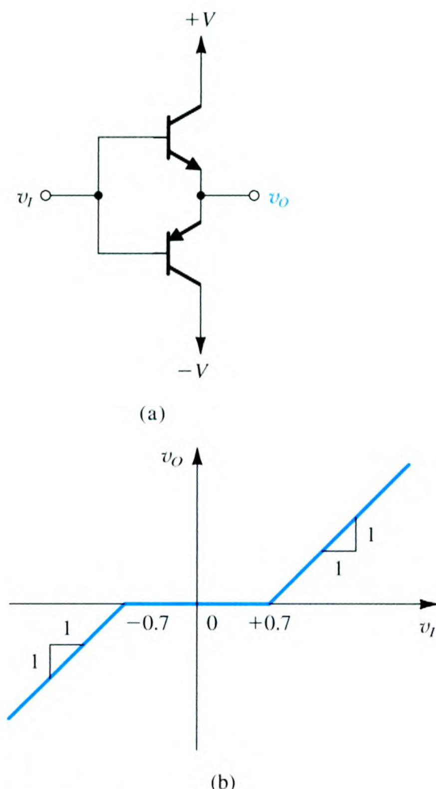


Figure P11.23

Section 11.3: The Feedback Voltage Amplifier

D 11.24 For the feedback voltage amplifier of Fig. 11.9(a), let the op amp have an infinite input resistance, a zero output resistance, and a finite open-loop gain of 1000 V/V. If $R_1 = 10\ \text{k}\Omega$, what value should R_2 have to obtain an ideal closed-loop gain of 10? Now, calculate the loop gain $A\beta$ and use it to find the actual value of the closed-loop gain A_f. If A_f is to be exactly 10, what must the value of R_2 be?

D 11.25 Consider the series–shunt feedback amplifier in Fig. 11.13(a), which is analyzed in Example 11.4.

(a) If $R_1 = 10\ \text{k}\Omega$, find the value of R_2 that results in an ideal closed-loop gain of 10.

(b) Use the expression for $A\beta$ derived in Example 11.4 to find the value of the loop gain for the case $\mu = 1000$, $R_{id} = 100\ \text{k}\Omega$, $r_o = 2\ \text{k}\Omega$, $R_s = 20\ \text{k}\Omega$, and $R_L = 10\ \text{k}\Omega$. Hence determine the value of the closed-loop gain A_f.

(c) By what factor must μ be increased to ensure that A_f is within 1% of the ideal value of 10?

D 11.26 Consider the series–shunt feedback amplifier of Fig. 11.9(b) that is analyzed in Example 11.3.

(a) If $R_1 = 1\ \text{k}\Omega$, what value must R_2 have to obtain a closed-loop gain whose ideal value is 5 V/V?

(b) If $g_{m1} = g_{m2} = 5$ mA/V, $R_{D1} = R_{D2} = 12\ \text{k}\Omega$, and the MOSFET's r_o is very large, use the expression for $A\beta$ derived in Example 11.3 to find the value of $A\beta$ and hence determine the closed-loop gain A_f.

D 11.27 Consider the series–shunt feedback amplifier of Fig. 11.9(c), which was the subject of Exercise 11.5. Assume that the voltage divider (R_1, R_2) is implemented with a 1-MΩ potentiometer. Assume that the MOSFET is biased so that $g_m = 5$ mA/V and r_o is large. Also, $R_D = 12\ \text{k}\Omega$. Find the value of R_1 that results in a closed-loop gain of 5 V/V.

D 11.28 VE 11.1 Figure P11.28 shows a series–shunt feedback amplifier known as a "feedback triple." All three MOSFETs are biased to operate at $g_m = 4$ mA/V. You may neglect their r_o's.

(a) Select a value for R_F that results in a closed-loop gain that is ideally 10 V/V.

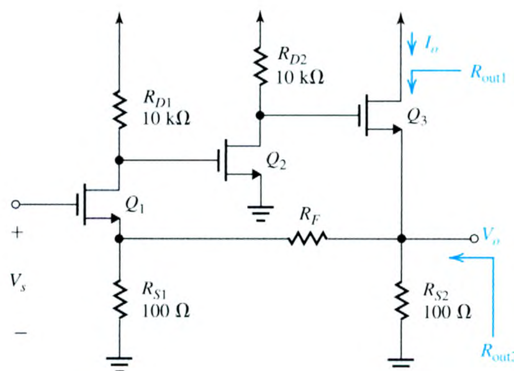


Figure P11.28

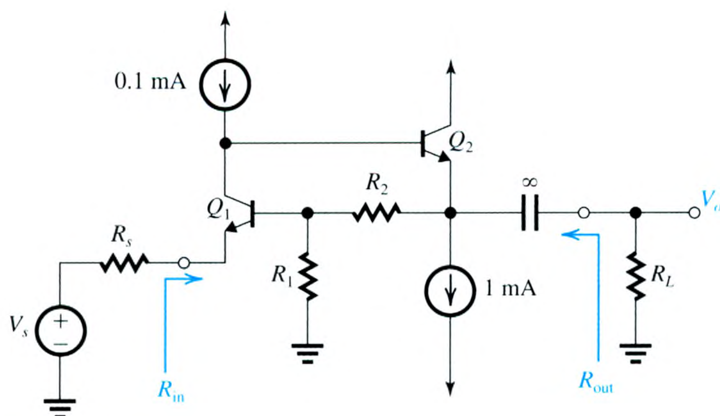


Figure P11.29

(b) Determine the loop gain $A\beta$ and hence the value of A_f. By what percentage does A_f differ from the ideal value you designed for? How can you adjust the circuit to make A_f equal to 10?

***11.29** In the series–shunt feedback amplifier shown in Fig. P11.29, the devices operate with $V_{BE} = 0.7$ V and have $\beta_1 = \beta_2 = 100$. The input signal V_s has a zero dc component. Resistances $R_s = 100\ \Omega$, $R_1 = 1\ \text{k}\Omega$, $R_2 = 10\ \text{k}\Omega$, and $R_L = 1\ \text{k}\Omega$.

(a) If the loop gain is large, what do you expect the closed-loop gain to be? Give both an expression and its value.
(b) Find the dc emitter current in each of Q_1 and Q_2. Also, find the dc voltage at the emitter of Q_2.
(c) Calculate the value of the loop gain $A\beta$. (*Hint*: Set $V_s = 0$ and break the loop at the base of Q_1. Simplify the circuit by eliminating dc sources.)
(d) Calculate the value of A_f.

11.30 Figure P11.30 shows a series–shunt feedback amplifier without details of the bias circuit.

(a) If R_E is selected to be 50 Ω, find the value for R_F that results in a closed-loop gain with an ideal value of 20 V/V.
(b) If Q_1 is biased at 0.5 mA, Q_2 at 1 mA, and Q_3 at 5 mA, and assuming that the transistors have $\beta = 100$ and large r_o, and that $R_{C1} = 2\ \text{k}\Omega$ and $R_{C2} = 1\ \text{k}\Omega$, find the value of the loop gain $A\beta$ and hence of the closed-loop gain A_f.

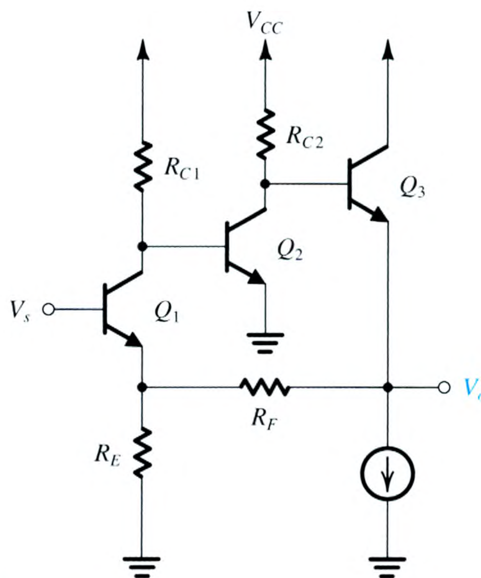


Figure P11.30

D 11.31 The current-mirror-loaded differential amplifier in Fig. P11.31 has a feedback network consisting of the voltage divider (R_1, R_2), with $R_1 + R_2 = 1\ \text{M}\Omega$. The devices are sized to operate at $|V_{OV}| = 0.2$ V. For all devices, $|V_A| = 10$ V. The input signal source has a zero dc component.

(a) Find the loop gain $A\beta$ and hence the value of A.
(b) Find the values of R_1 and R_2 that result in a closed-loop gain of exactly 5 V/V.

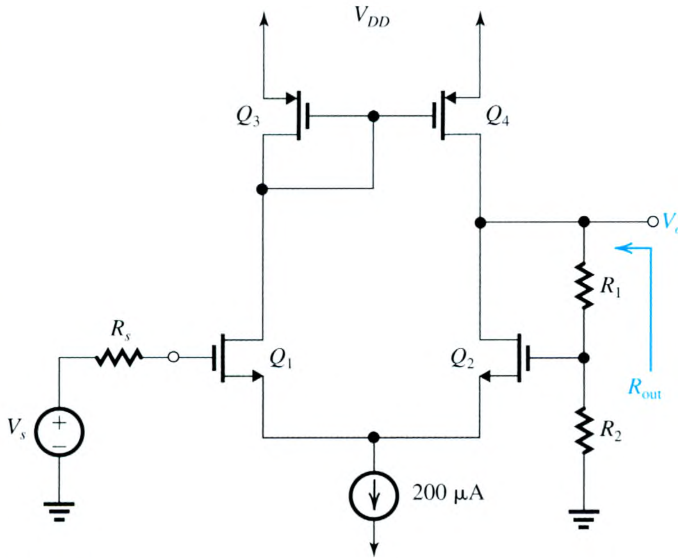


Figure P11.31

Section 11.4: Systematic Analysis of Feedback Voltage Amplifiers (Series–Shunt)

11.32 A series–shunt feedback amplifier employs a basic amplifier with input and output resistances each of 2 kΩ and gain $A = 2000$ V/V. The feedback factor $\beta = 0.1$ V/V. Find the gain A_f, the input resistance R_{if}, and the output resistance R_{of} of the closed-loop amplifier.

11.33 For a particular amplifier connected in a feedback loop in which the output voltage is sampled, measurement of the output resistance before and after the loop is connected shows a change by a factor of 100. Is the resistance with feedback higher or lower? What is the value of the loop gain $A\beta$? If R_{of} is 100 Ω, what is R_o without feedback?

11.34 The formulas for R_{if} and R_{of} in Eqs. (11.22) and (11.25), respectively, also apply for the case in which A is a function of frequency. In this case, the resulting impedances Z_{if} and Z_{of} will be functions of frequency. Consider the case of a series–shunt amplifier that has an input resistance R_i, an output resistance R_o, and open-loop gain $A = A_0/[1 + (s/\omega_H)]$, and a feedback factor β that is independent of frequency. Find Z_{if} and Z_{of} and give an equivalent circuit for each, together with the values of all the elements in the equivalent circuits.

11.35 A feedback amplifier utilizing voltage sampling and employing a basic voltage amplifier with a gain of 1000 V/V and an input resistance of 1000 Ω has a closed-loop input resistance of 100 kΩ. What is the closed-loop gain? If the basic amplifier is used to implement a unity-gain voltage buffer, what input resistance do you expect?

11.36 Consider the noninverting op-amp circuit of Example 11.5 for the case $R_1 = \infty$ and $R_2 = 0$.

(a) What is the value of β, and what is the ideal value of the closed-loop gain?
(b) Adapt the expressions found in Example 11.5 to obtain expressions for A and $A\beta$ for this case.
(c) For $\mu = 10^4$, $R_{id} = 100$ kΩ, $R_s = 10$ kΩ, $r_o = 1$ kΩ, and $R_L = 2$ kΩ, find A, $A\beta$, A_f, R_{in}, and R_{out}.

SIM **D *11.37** Figure P11.37 shows a series–shunt amplifier with a feedback factor $\beta = 1$. The amplifier is designed so that $v_O = 0$ for $v_S = 0$, with small deviations in v_O from 0 V dc being minimized by the negative-feedback action. The technology utilized has $k_n' = 4k_p' = 400\ \mu A/V^2$, $|V_t| = 0.35$ V, and $|V_A'| = 10$ V/μm.

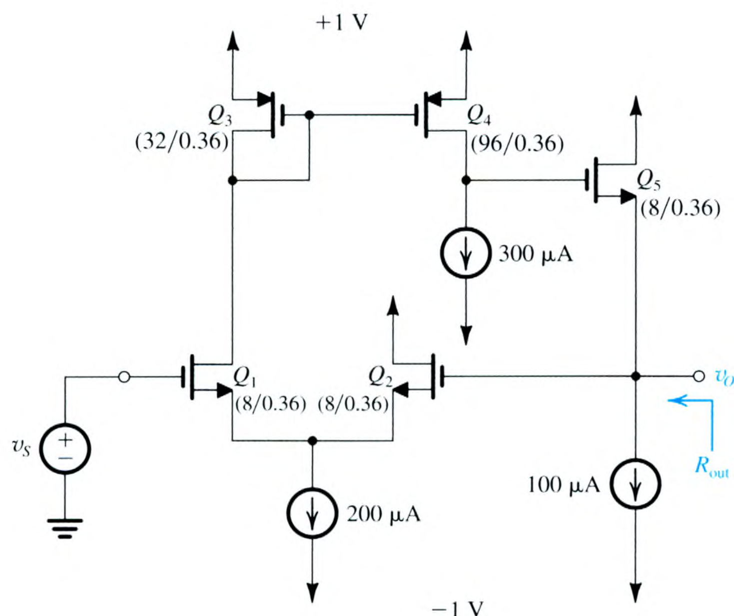


Figure P11.37

(a) Show that the feedback is negative.

(b) With the feedback loop opened at the gate of Q_2, and the gate terminals of Q_1 and Q_2 grounded, find the dc current and the overdrive voltage at which each of Q_1 to Q_5 is operating. Ignore the Early effect. Also find the dc voltage at the output.

(c) Find g_m and r_o of each of the five transistors.

(d) Find expressions and values of A and R_o. Assume that the bias current sources are ideal.

(e) Find the gain with feedback, A_f, and the output resistance R_{out}.

(f) How would you modify the circuit to realize a closed-loop voltage gain of 5 V/V? What is the value of output resistance obtained?

***11.38** This problem deals with the series–shunt feedback amplifier of Fig. P11.29 and overlaps somewhat with Problem 11.29. Thus, if you have already solved 11.29, you can use some of the results in the solution of this problem. The devices operate with $V_{BE} = 0.7$ V and have $\beta_1 = \beta_2 = 100$. The input signal V_s has a zero dc component. Resistances $R_s = 100\ \Omega$, $R_1 = 1\ \text{k}\Omega$, $R_2 = 10\ \text{k}\Omega$, and $R_L = 1\ \text{k}\Omega$.

(a) If the loop gain is large, what do you expect the closed-loop gain V_o/V_s to be? Give both an expression and its approximate value.

(b) Find the dc emitter current in each of Q_1 and Q_2. Also find the dc voltage at the emitter of Q_2.

(c) Sketch the A circuit without the dc sources. Derive expressions for A, R_i, and R_o, and find their values.

(d) Give an expression for β and find its value.

(e) Find the closed-loop gain V_o/V_s, the input resistance R_{in}, and the output resistance R_{out}. By what percentage does the value of A_f differ from the approximate value found in (a)?

***11.39** Figure P11.39 shows a series–shunt amplifier in which the three MOSFETs are sized to operate at $|V_{OV}| = 0.2$ V. Let $|V_t| = 0.5$ V and $|V_A| = 10$ V. The current sources utilize single transistors and thus have output resistances equal to r_o.

(a) Show that the feedback is negative.

(b) Assuming the loop gain to be large, what do you expect the closed-loop voltage gain V_o/V_s to be approximately?

(c) If V_s has a zero dc component, find the dc voltages at nodes S_1, G_2, S_3, and G_3. Verify that each of the current sources has the minimum required dc voltage across it for proper operation.

(d) Find the A circuit. Calculate the gain of each of the three stages and the overall voltage gain, A.

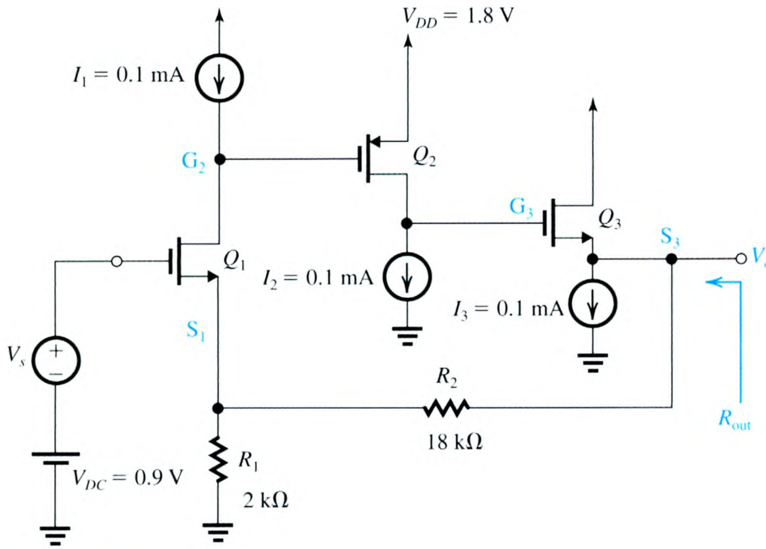


Figure P11.39

[*Hint*: A CS amplifier with a resistance R_s in the source lead has an effective transconductance $g_m/(1 + g_m R_s)$ and an output resistance $r_o(1 + g_m R_s)$.]

(e) Find β.

(f) Find $A_f = V_o/V_s$. By what percentage does this value differ from the approximate value obtained in (b)?

(g) Find the output resistance R_{out}.

D *11.40 This problem deals with the series–shunt feedback amplifier of Fig. P11.31. Certain aspects of this amplifier were considered in Problem 11.31. If you have already solved problem 11.31, you will have the opportunity to compare results. The current-mirror-loaded differential amplifier has a feedback network consisting of the voltage divider (R_1, R_2), with $R_1 + R_2 = 1$ MΩ. The devices are sized to operate at $|V_{OV}| = 0.2$ V. For all devices, $|V_A| = 10$ V. The input signal source has a zero dc component.

VE 11.2

(a) Show that the feedback is negative.

(b) What do you expect the dc voltage at the gate of Q_2 to be? At the output? (Neglect the Early effect.)

(c) Find the *A* circuit. Derive an expression for *A* and find its value.

(d) Select values for R_1 and R_2 to obtain a closed-loop voltage gain $V_o/V_s = 5$ V/V.

(e) Find the value of R_{out}.

(f) Utilizing the open-circuit closed-loop gain (5 V/V) and the value of R_{out} found in (e), find the value of gain obtained when a resistance $R_L = 10$ kΩ is connected to the output.

(g) As an alternative approach to (f) above, redo the analysis of the *A* circuit including R_L. Then utilize the values of R_1 and R_2 found in (d) to determine β and A_f. Compare the value of A_f to that found in (f).

D *11.41 Figure P11.30 shows a series–shunt feedback amplifier without details of the bias circuit.

(a) Eliminating the dc sources, sketch the *A* circuit and the circuit for determining β.

(b) Show that the ideal value of the closed-loop voltage gain is given by

$$A_f \equiv \frac{V_o}{V_s} = \frac{R_F + R_E}{R_E}$$

(c) If R_E is selected equal to 50 Ω, find R_F that will result in a closed-loop gain of approximately 20 V/V.

(d) If Q_1 is biased at 0.5 mA, Q_2 at 1 mA, and Q_3 at 5 mA, and assuming that the transistors have $\beta = 100$, find approximate values for R_{C1} and R_{C2} to obtain gains from

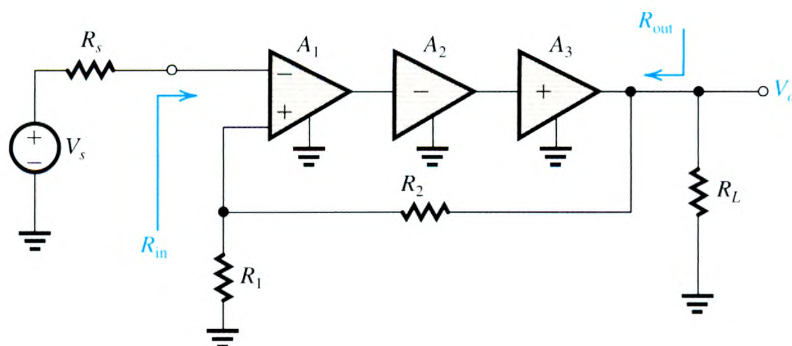


Figure P11.42

the stages of the A circuit as follows: a voltage gain of Q_1 of about –10 V/V and a voltage gain of Q_2 of about –50 V/V.

(e) For your design, what is the closed-loop voltage gain realized?

(f) Calculate the input and output resistances of the closed-loop amplifier designed.

D *11.42 Figure P11.42 shows a three-stage feedback amplifier:

A_1 has an 82-kΩ differential input resistance, a 20-V/V open-circuit differential voltage gain, and a 3.2-kΩ output resistance.

A_2 has a 5-kΩ input resistance, a 20-mA/V short-circuit transconductance, and a 20-kΩ output resistance.

A_3 has a 20-kΩ input resistance, unity open-circuit voltage gain, and a 1-kΩ output resistance.

The feedback amplifier feeds a 1-kΩ load resistance and is fed by a signal source with a 9-kΩ resistance.

(a) Show that the feedback is negative.

(b) If $R_1 = 20$ kΩ, find the value of R_2 that results in a closed-loop gain V_o/V_s that is ideally 5 V/V.

(c) Supply the small-signal equivalent circuit.

(d) Sketch the A circuit and determine A.

(e) Find β and the amount of feedback.

(f) Find the closed-loop gain $A_f \equiv V_o/V_s$.

(g) Find the feedback amplifier's input resistance R_{in}.

(h) Find the feedback amplifier's output resistance R_{out}.

(i) If the high-frequency response of the open-loop gain A is dominated by a pole at 100 Hz, what is the upper 3-dB frequency of the closed-loop gain?

(j) If for some reason A_1 drops to half its nominal value, what is the percentage change in A_f?

Section 11.5: Other Feedback-Amplifier Types

D 11.43 Negative feedback is to be used to modify the characteristics of a particular amplifier for various purposes. Identify the feedback topology to be used if:

(a) input resistance is to be lowered and output resistance raised.

(b) both input and output resistances are to be raised.

(c) both input and output resistances are to be lowered.

11.44 A feedback transconductance amplifier is to have an ideal closed-loop gain of 10 mA/V. Find the required value of β. If the open-loop transconductance amplifier has a gain of 1 A/V, an input resistance of 10 kΩ, and an output resistance of 100 kΩ, find the actual closed-loop gain, the input resistance, and the output resistance of the feedback amplifier.

11.45 Figure P11.45(a) shows a feedback transconductance amplifier using an op amp together with an NMOS transistor. The feedback network consists of a resistor R_F that senses the output current I_o (recall that the drain and source currents of a MOSFET are equal) and provides a feedback voltage that is subtracted from V_s by means of the differencing action of the op-amp input.

(a) Setting $A = \infty$, find the ideal closed-loop gain I_o/V_s.

(b) Find the value of R_F that results in an ideal closed-loop gain of 10 mA/V.

(c) Find an expression for the loop gain $A\beta$ by setting $V_s = 0$, breaking the loop, and applying an input test voltage V_t, as shown in Fig. P11.45(b). Assume the input resistance of the op amp to be very high.

(d) For $\mu = 1000$, and assuming the MOSFET has $g_m = 2$ mA/V and $r_o = 20$ kΩ, find the values of $A\beta$, A, and A_f.

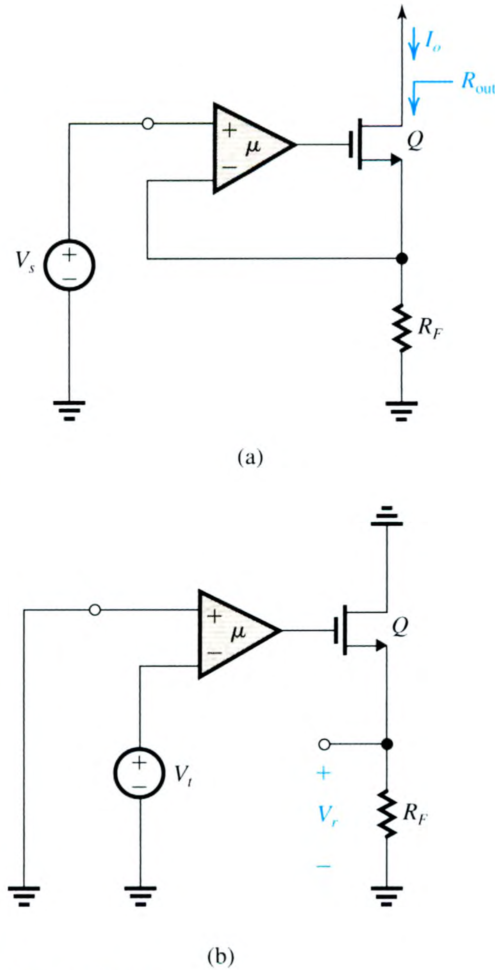


Figure P11.45

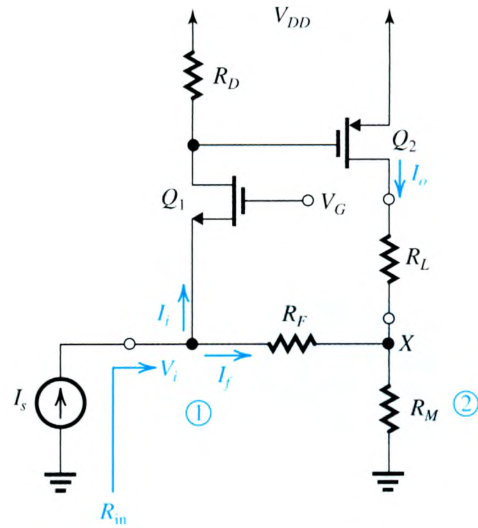


Figure P11.46

D 11.46 Figure P11.46 shows a feedback current amplifier. The feedback network consists of the highlighted two-port network comprising R_M and R_F. It is fed with the output current I_o and delivers a feedback current I_f at its port 1 to the input node.

(a) By setting the open-loop gain to ∞, which results in $I_i = 0$ and hence $V_i = 0$, show that the ideal closed-loop gain is

$$A_f\,|_{\text{ideal}} \equiv \frac{I_o}{I_s} = -\left(1 + \frac{R_F}{R_M}\right)$$

Selecting $R_F = 1\ \text{k}\Omega$, find R_M that results in an ideal closed-loop gain of -5 A/A.

(b) Find an expression for β.

(c) Setting $I_s = 0$, break the loop at the gate of Q_2 and thus determine the loop gain $A\beta$. Neglecting r_o of each of Q_1 and Q_2, show that

$$A = -\frac{g_{m2}R_D}{1 + 1/[g_{m1}(R_M + R_F)]}$$

(d) Find an expression for the closed-loop gain A_f.

(e) For $g_{m1} = g_{m2} = 4$ mA/V, $R_D = 10\ \text{k}\Omega$, and $R_F = 1\ \text{k}\Omega$, find the value of R_M that results in a closed-loop gain of -5 A/A. Compare to the value found in (a).

D 11.47 A basic transresistance amplifier has an open-loop gain $A \equiv V_o/I_s$ of 100 V/mA, an input resistance of 1 kΩ, and an output resistance of 1 kΩ. What feedback topology is required to reduce both input and output resistance to 10 Ω? What is the value of β required to achieve this? What is the resulting closed-loop gain?

Feedback Transconductance Amplifiers (Series–Series)

11.48 A series–series feedback amplifier employs a transconductance amplifier having a short-circuit transconductance G_m of 0.6 A/V, input resistance of 10 kΩ, and output resistance of 100 kΩ. The feedback network has β = 200 Ω, an input resistance (with port 1 open-circuited) of 200 Ω, and an input resistance (with port 2 open-circuited) of 10 kΩ. The amplifier operates with a signal source having a resistance of 10 kΩ and with a load resistance of 10 kΩ. Find A_f, R_{in}, and R_{out}.

VE 11.3 **11.49** The feedback transconductance amplifier in Fig. P11.45(a) uses a differential voltage amplifier with a gain μ and a very high input resistance, and a MOSFET Q having $g_m = 5$ mA/V and $r_o = 20$ kΩ. Design the circuit to obtain an ideal closed-loop transconductance of 10 mA/V and an output resistance R_{out} of 5 MΩ. What are the values required for R_F and μ? Also, what is the actual value obtained of the closed-loop transconductance?

11.50 Reconsider the circuit in Fig. 11.24(a), analyzed in Example 11.8, this time with the output voltage taken at the emitter of Q_3. In this case the feedback can be considered to be of the series–shunt type. Note that R_{E2} should now be considered part of the basic amplifier and not of the feedback network.

(a) Find expressions for $A_f|_{ideal} \equiv V_{e3}/V_s$ and β. Also, find the values of both.
(b) If the loop gain remains at the value calculated in Example 11.8 (that is, 246.3), find the actual value realized of the closed-loop voltage gain V_{e3}/V_s. Also, find the value of the output resistance R_{out} measured between the emitter of Q_3 and ground. Neglect the effect of r_{o3}.
Note: If you analyze this circuit from scratch, you will find that the loop gain is slightly different than the value found in Example 11.8. This is due to the different approximations inherent in the feedback analysis method.

11.51 Figure P11.51 shows a feedback transconductance amplifier composed of a differential voltage amplifier A_1 with an input resistance R_{id}, an open-circuit voltage gain A_1, and an output resistance r_{o1}, connected in cascade with a common-source MOSFET Q_2 having a transconductance g_{m2} and an output resistance r_{o2}.

(a) Setting $A = \infty$, find an expression for $A_f|_{ideal} = I_o/V_s$.
(b) Find an expression for β and give the β circuit.
(c) Find the loading effects of the β circuit; that is, resistances R_{11} and R_{22}.
(d) Give the A circuit with A_1 and Q_2 replaced with their equivalent circuits.
(e) Find an expression for A. Simplify the expression for the case $R_{id} \gg (R_s + R_F)$ and $r_{o2} \gg (R_L + R_F)$.
(f) Use the simplified expression for A to determine A_f, R_{in}, and R_{out}.
(g) For $R_F = 200$ Ω, $A_1 = 200$ V/V, $R_{id} = 100$ kΩ, $g_{m2} = 2$ mA/V, $r_{o2} = 20$ kΩ, and $R_s = R_L = 1$ kΩ, find the values of A_f, R_{in}, and R_{out}.

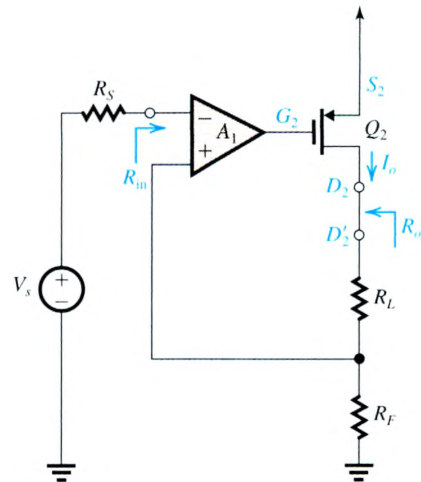


Figure P11.51

D 11.52 Figure P11.28 shows a feedback triple utilizing MOSFETs. All three MOSFETs are biased and sized to operate at $g_m = 5$ mA/V. You may neglect their r_o's (except for the calculation of R_{out1}, as indicated below).

(a) Setting $A = \infty$, derive an expression for the ideal closed-loop gain $A_f|_{ideal} = I_o/V_s$. Use this expression to determine the value of R_F that results in $A_f|_{ideal} = 100$ mA/V.
(b) Give the β circuit, and find the value of β and the two loading resistances R_{11} and R_{22}.
(c) Sketch the A circuit and find the value of A.
(d) Find $(1 + A\beta)$ and A_f. Compare to the value of A_f you designed for. What is the percentage difference? What resistance can you change to make A_f exactly 100 mA/V, and in which direction (increase or decrease)?
(e) Assuming $r_{o3} = 20$ kΩ, find R_o of the A circuit. [Recall that the resistance looking into the drain of a MOSFET having a resistance R_s in its source is $(r_o + R_s + g_m r_o R_s)$.] Hence, find the output resistance of the feedback transconductance amplifier R_{out1}. Since the current sampled by the feedback network is exactly equal to the output current, you can use the feedback formula.

D 11.53 The feedback transconductance amplifier in Fig. P11.53 uses a differential amplifier with a voltage gain μ, an input resistance $R_{id} = 100$ kΩ, and an output resistance $r_{o1} = 1$ kΩ; and a MOSFET with $g_m = 2$ mA/V and $r_{o2} = 20$ kΩ.

(a) Find the value of R_2 that results in the closed-loop gain I_o/V_s having an ideal value of 100 mA/V.
(b) Find the value of β and sketch the two-port feedback circuit.

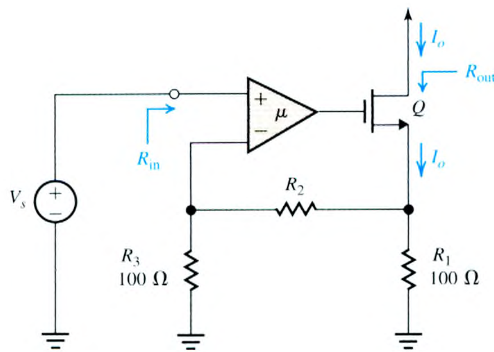


Figure P11.53

(c) Find the loading effects of the feedback circuit, R_{11} and R_{22}.

(d) Give the A circuit and find an expression for A in terms of μ.

(e) Find the value of μ that results in a 40-dB amount of feedback.

(f) Find the realized value of A_f.

(g) Find R_{in} and R_{out}.

11.54 Consider the circuit in Fig. P11.54 as a transconductance amplifier with input V_s and output I_o. The transistor is specified in terms of its g_m and r_o.

(a) Sketch the small-signal equivalent circuit using the hybrid-π model of the MOSFET and convince yourself that the feedback circuit is comprised of resistor R_F.

(b) Find the β circuit and the A circuit.

(c) Derive expressions for A, β, $(1+A\beta)$, A_f, R_o, and R_{of}.

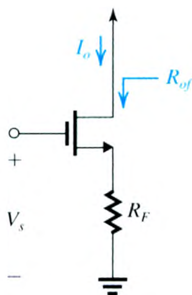


Figure P11.54

***11.55** Show that the output resistance of the BJT circuit in Fig. P11.55 is given by

$$R_o = r_o + [R_e \| (r_\pi + R_b)] \left(1 + g_m r_o \frac{r_\pi}{r_\pi + R_b}\right)$$

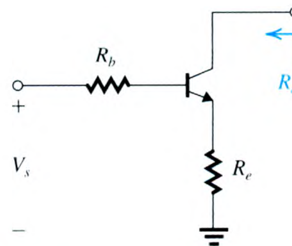
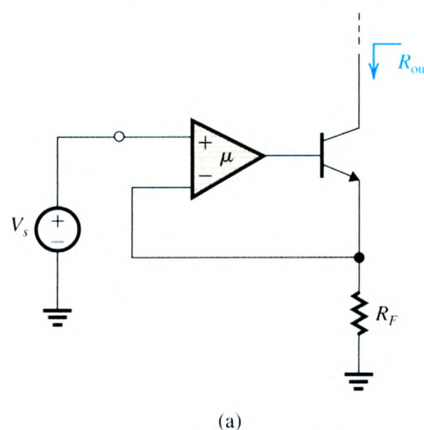


Figure P11.55

To derive this expression, set $V_s = 0$, replace the BJT with its small-signal, hybrid-π model, apply a test voltage V_x to the collector, and find the current I_x drawn from V_x and hence R_o as V_x/I_x. Note that the bias arrangement is not shown. For the case of $R_b = 0$, find the maximum possible value for R_o. Note that this theoretical maximum is obtained when R_e is so large that the signal current in the emitter is nearly zero. In this case, with V_x applied and $V_s = 0$, what is the current in the base, in the $g_m V_\pi$ generator, and in r_o, all in terms of I_x? Show these currents on a sketch of the equivalent circuit with R_e set to ∞.

11.56 As we found out in Example 11.8, whenever the feedback network senses the emitter current of the BJT, the feedback output resistance formula cannot predict the output resistance looking into the collector. To understand this issue more clearly, consider the feedback transconductance amplifier shown in Fig. P11.56(a). To determine the output resistance, we set $V_s = 0$ and apply a test voltage V_x to the collector, as shown in Fig. P11.56(b). Now, let μ be increased to the point where the feedback signal across R_F almost equals the input to the positive terminal of the differential amplifier,



(a)

Figure P11.56

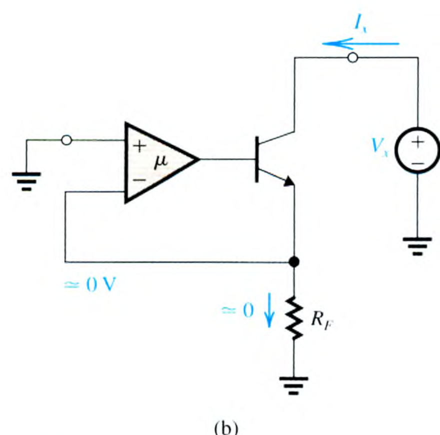


Figure P11.56 *continued*

now zero. Thus the signal current through R_F will be almost zero. By replacing the BJT with its hybrid-π model, show that

$$R_{\text{out}} = r_\pi + (h_{fe} + 1)r_o \simeq h_{fe}r_o$$

where h_{fe} is the transistor β. Thus for large amounts of feedback, R_{out} is limited to a maximum of $h_{fe}r_o$ independent of the amount of feedback. This phenomenon does *not* occur in the MOSFET version of this circuit, where the output resistance can be theoretically made infinite.

11.57 For the feedback transconductance amplifier of Fig. P11.57 derive expressions for A, β, $A\beta$, A_f, R_o, and R_{of}. Evaluate A_f and R_{of} for the case of $g_{m1} = g_{m2} = 4$ mA/V, $R_D = 25$ kΩ, $r_{o2} = 20$ kΩ, $R_F = 100$ Ω, and $R_L = 1$ kΩ. For simplicity, neglect r_{o1} and take r_{o2} into account only when calculating output resistances. Also, evaluate the output resistance R_{out} seen between the two terminals of R_L.

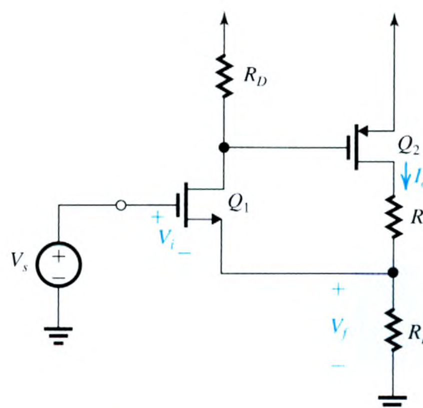


Figure P11.57

SIM **11.58** All the MOS transistors in the feedback transconductance amplifier (series–series) of Fig. P11.58 are

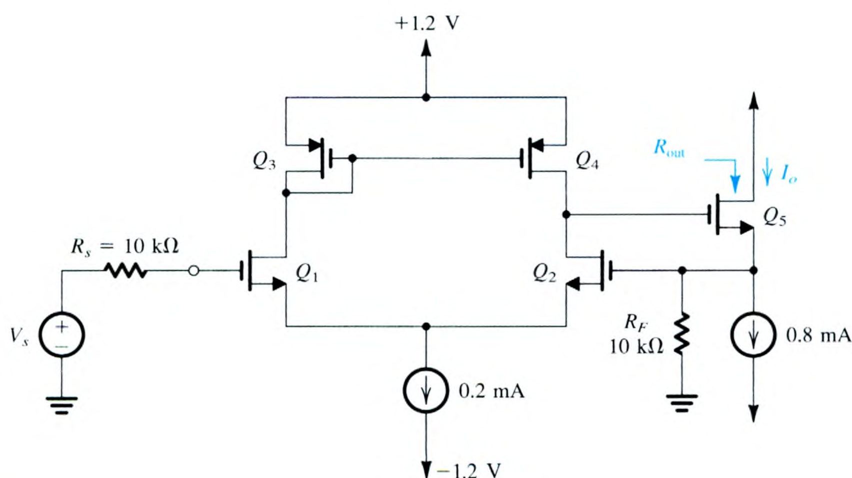


Figure P11.58

sized to operate at $|V_{OV}| = 0.2$ V. For all transistors, $|V_t| = 0.4$ V and $|V_A| = 20$ V.

(a) If V_s has a zero dc component, find the dc voltage at the gate of Q_2, at the drain of Q_1, and at the drain of Q_2.
(b) Find an expression and value for $A_f \equiv I_o/V_s$ for the case $A\beta \gg 1$.
(c) Use feedback analysis to obtain a more precise value for A_f.
(d) Find the value of R_{out}.
(e) If the voltage at the source of Q_5 is taken as the output, find the voltage gain using the value of I_o/V_s obtained in (c). Also find the output resistance of this series–shunt voltage amplifier.

11.59 By setting $V_s = 0$ and breaking the feedback loop, show that the loop gain of the amplifier circuit in Fig. P11.58 is

$$A\beta = g_{m1,2}(r_{o2} \parallel r_{o4}) \frac{R_F \parallel r_{o5}}{(R_F \parallel r_{o5}) + 1/g_{m5}}$$

where $g_{m1,2}$ is the g_m of each of Q_1 and Q_2.

Feedback Transresistance Amplifiers (Shunt–Shunt)

11.60 A shunt–shunt feedback amplifier uses a transresistance amplifier with an open-circuit transresistance of 100 V/mA, input resistance of 1 kΩ, and output resistance of 1 kΩ. The feedback network has $\beta = 1$ mA/V, input resistance at port 1 with port 2 short-circuited of 1 kΩ, and input resistance at port 2 with port 1 short-circuited of 1 kΩ. The amplifier is fed with a signal current source having a resistance of 10 kΩ, and is connected to a load of 1-kΩ resistance. Find the closed-loop transresistance A_f, the input resistance R_{in}, and the output resistance, R_{out}.

11.61 By setting $I_s = 0$, breaking the feedback loop, and replacing the MOSFET with its hybrid-π model, determine the loop gain of the feedback amplifier in Fig. E11.15. Hence, find the open-loop gain. Evaluate $A\beta$, β, A, and A_f for the numerical values given in Exercise 11.15. Why do the results differ somewhat from those given in the answer to Exercise 11.15?

D 11.62 The circuit in Fig. P11.62 utilizes a voltage amplifier with gain μ in a shunt–shunt feedback topology with the feedback network composed of resistor R_F. In order to be able to use the feedback equations, you should first convert the signal source to its Norton representation. You will then see that the analysis will be similar to that used in Example 11.9.

(a) If the loop gain is very large, what approximate closed-loop voltage gain V_o/V_s is realized? If $R_s = 10$ kΩ, give the value of R_F that will result in $V_o/V_s \simeq -10$ V/V.
(b) If the amplifier μ has a dc gain of 10^3 V/V, an input resistance $R_{id} = 1$ MΩ, and an output resistance $r_o = 1$ kΩ, find the actual V_o/V_s realized. Also find R_{in} and R_{out} (indicated on the circuit diagram). You may use formulas derived in Example 11.9.
(c) If the amplifier μ has an upper 3-dB frequency of 2 kHz and a uniform −20-dB/decade gain rolloff, what is the 3-dB frequency of the gain $|V_o/V_s|$?

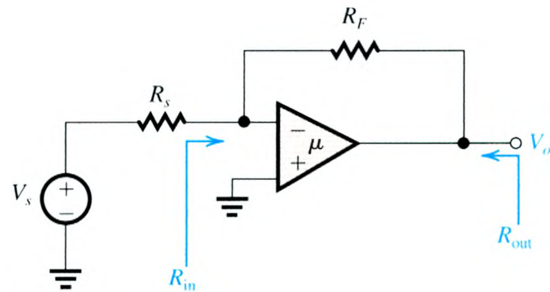


Figure P11.62

11.63 The CE BJT amplifier in Fig. P11.63 employs shunt–shunt feedback: Feedback resistor R_F senses the output voltage V_o and provides a feedback current to the base node.

(a) If V_s has a zero dc component, find the dc collector current of the BJT. Assume the transistor $\beta = 100$ and $V_{BE} = 0.7$ V.
(b) Find the small-signal equivalent circuit of the amplifier with the signal source represented by its Norton equivalent (as we usually do when the feedback connection at the input is shunt). Neglect r_o.
(c) Find the A circuit and determine the value of A, R_i, and R_o.
(d) Find β and hence $A\beta$ and $1 + A\beta$.
(e) Find A_f, R_{if}, and R_{of} and hence R_{in} and R_{out}.

(f) What voltage gain V_o/V_s is realized? How does this value compare to the ideal value obtained if the loop gain is very large and thus the signal voltage at the base becomes almost zero (like what happens in an inverting op-amp circuit). Note that this single-transistor poor-man's op amp is not that bad!

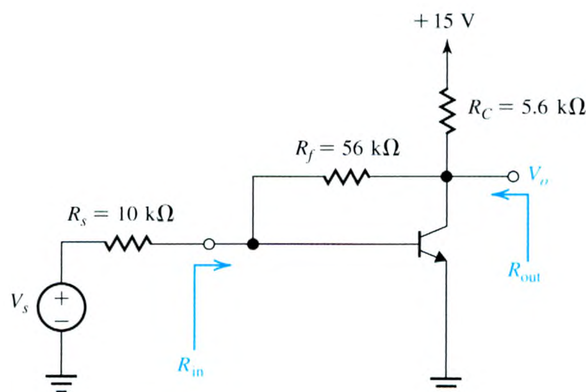


Figure P11.63

11.64 Analyze the circuit in Fig. E11.15 from first principles (i.e., do not use the feedback approach) and hence show that

$$A_f \equiv \frac{V_o}{I_s} = -\frac{(R_s \| R_F)\left(g_m - \dfrac{1}{R_F}\right)(r_o \| R_F)}{1 + (R_s \| R_F)\left(g_m - \dfrac{1}{R_F}\right)(r_o \| R_F)/R_F}$$

Comparing this expression to the one given in Exercise 11.15, part (b), you will note that the only difference is that g_m has been replaced by $(g_m - 1/R_F)$. Note that $-1/R_F$ represents the forward transmission in the feedback network, which the feedback-analysis method neglects. What is the condition then for the feedback-analysis method to be reasonably accurate for this circuit?

11.65 The feedback transresistance amplifier in Fig. P11.65 utilizes two identical MOSFETs biased by ideal current sources $I = 0.3$ mA. The MOSFETs are sized to operate at $V_{OV} = 0.15$ V and have $V_t = 0.35$ V and $V_A = 6$ V. The feedback resistance $R_F = 20$ kΩ.

(a) If I_s has a zero dc component, find the dc voltage at the input, at the drain of Q_1, and at the output.
(b) Find g_m and r_o of Q_1 and Q_2.
(c) Replacing the MOSFETs with their equivalent circuit models, provide the A circuit and derive an expression for A in terms of g_{m1}, r_{o1}, g_{m2}, r_{o2}, and R_F.
(d) What is β? Give an expression for the loop gain $A\beta$ and the amount of feedback $(1 + A\beta)$.
(e) Derive an expression for A_f.
(f) Derive expressions for R_i, R_{in}, R_o, and R_{out}.
(g) Evaluate A, β, $A\beta$, A_f, R_i, R_o, R_{in}, and R_{out} for the component values given.

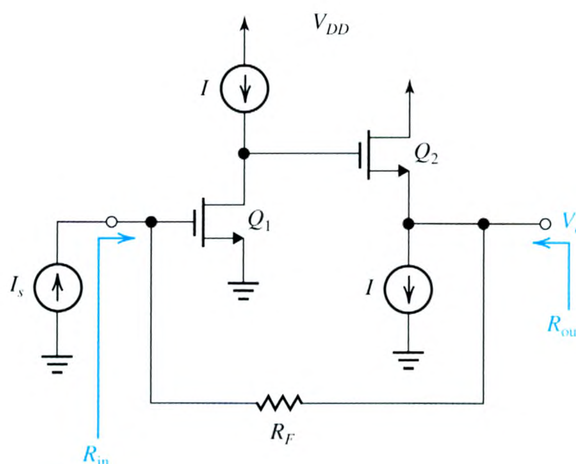


Figure P11.65

11.66 By setting $I_s = 0$ and breaking the feedback loop, find the loop gain of the feedback amplifier in Fig. P11.65. If you have already solved Problem 11.65, compare results. Which result do you think is more accurate, and why? For the numerical values given in Problem 11.65, by how much (in percent) do the two values of loop gain differ?

11.67 For the feedback transresistance amplifier in Fig. P11.67, let $V_{CC} = -V_{EE} = 5$ V, $R_C = 7$ kΩ, $R_E = 11.5$ kΩ, and $R_F = 10$ kΩ. The transistors have $V_{BE} = 0.7$ V and $\beta = 100$.

(a) If I_s has a zero dc component, show that Q_1 and Q_2 are operating at equal dc collector currents of approximately 0.5 mA. What is the dc voltage at the output?
(b) Replacing each of Q_1 and Q_2 with their equivalent circuit models, find the A circuit and the value of A, R_i, and R_o. Neglect r_{o1} and r_{o2}.
(c) Find the value of β, the loop gain, and the amount of feedback.

(d) Find $A_f \equiv V_o/I_s$, the input resistance R_{in}, and the output resistance R_{out}.

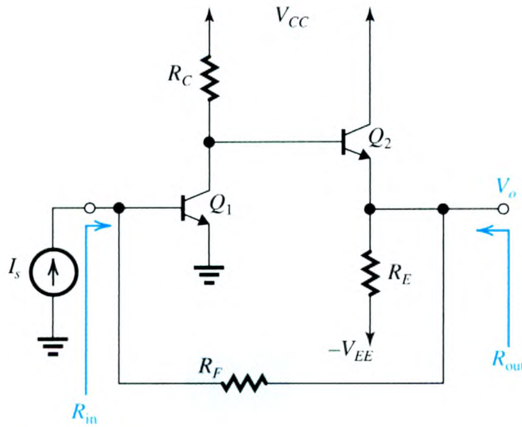


Figure P11.67

D 11.68 For the feedback amplifier in Fig. P11.68, select a value for R_F that results in an ideal closed-loop gain $A_f \equiv V_o/I_s = -20$ V/mA. Then, analyze the circuit to determine the actual value of A_f realized. As well, determine R_{in} and R_{out}. Transistors Q_1 and Q_2 are operated so that $g_{m1} = g_{m2} = 2$ mA/V and r_{o1} and r_{o2} can be neglected. Also, $R_{D1} = R_{D2} = 20$ kΩ.

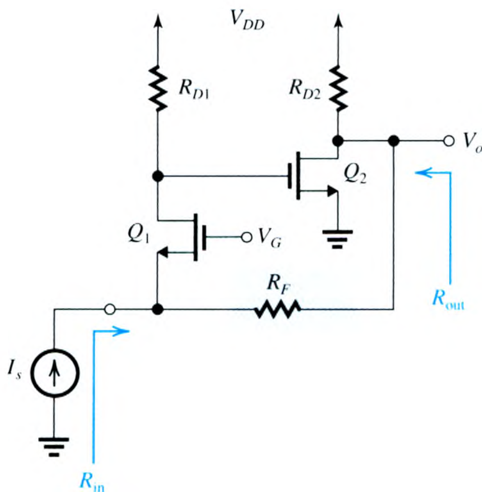


Figure P11.68

11.69 The feedback amplifier of Fig. P11.69 consists of a common-gate amplifier formed by Q_1 and R_D, and a feedback circuit formed by the capacitive divider (C_1, C_2) and the common-source transistor Q_f. Note that the bias circuit for Q_f is not shown. It is required to derive expressions for $A_f \equiv V_o/I_s$, R_{in}, and R_{out}. Assume that C_1 and C_2 are sufficiently small that their loading effect on the basic amplifier can be neglected. Also neglect r_o. Find the values of A_f, R_{in}, and R_{out} for the case in which $g_{m1} = 5$ mA/V, $R_D = 10$ kΩ, $C_1 = 0.9$ pF, $C_2 = 0.1$ pF, and $g_{mf} = 2$ mA/V.

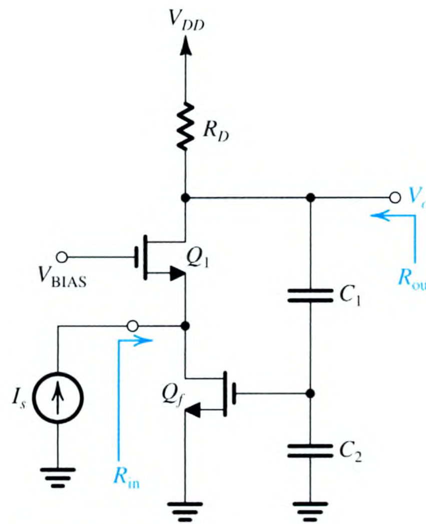


Figure P11.69

D *11.70 Figure P11.70 shows a shunt–shunt feedback amplifier. The MOSFETs have $V_{tn} = 0.4$ V, $V_A = 10$ V, and $\mu_n C_{ox} = 400$ μA/V^2. The power supply $V_{DD} = 1.8$ V, and $R_L = 10$ kΩ. The coupling capacitor C_C can be assumed to be very large.

(a) Perform a dc design to meet the following specifications: $I_{D1} = I_{D2} = 100$ μA, $I_{R2,R1} = 10$ μA, $V_{OV1} = V_{OV2} = 0.2$ V. Neglect the Early effect. Specify the values required for I_1, R_1, R_2, $(W/L)_1$, and $(W/L)_2$.
(b) Find expressions for the ideal value of V_o/V_s and for β.
(c) Find the value of R_s that results in V_o/V_s being ideally -5 V/V.
(d) Find the A circuit and use it to determine the values of A, R_i, and R_o.
(e) Find the value obtained for V_o/V_s.
(f) Find R_{in} and R_{out}.

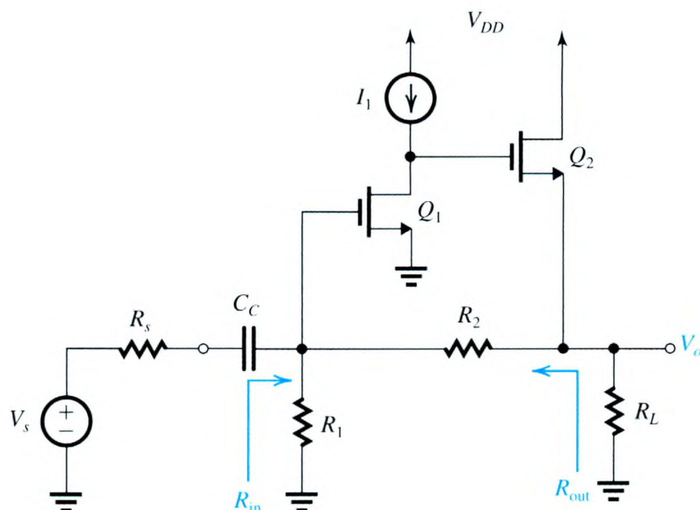
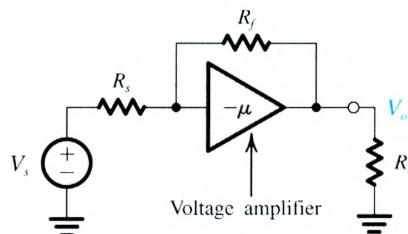


Figure P11.70

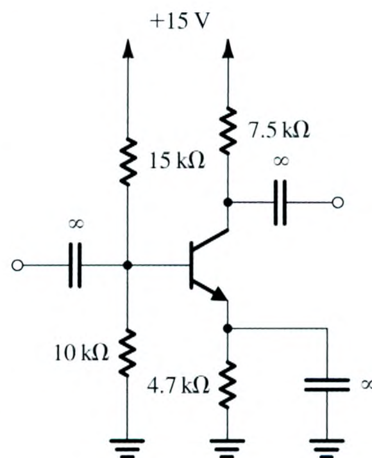
D **11.71 (a) Show that for the circuit in Fig. P11.71(a), if the loop gain is large, the voltage gain V_o/V_s is given approximately by

$$\frac{V_o}{V_s} \simeq -\frac{R_f}{R_s}$$

(b) Using three cascaded stages of the type shown in Fig. P11.71(b) to implement the amplifier μ, design a feedback amplifier with a voltage gain of approximately −100 V/V. The amplifier is to operate between a source resistance $R_s = 10\ \text{k}\Omega$ and a load resistance $R_L = 1\ \text{k}\Omega$. Calculate the actual value of V_o/V_s realized, the input resistance (excluding R_s), and the output resistance (excluding R_L). Assume that the BJTs have h_{fe} of 100. (*Note:* In practice, the three amplifier stages are not made identical, for stability reasons.)



(a)



(b)

Figure P11.71

Feedback Current Amplifiers (Shunt–Series)

D 11.72 Design the feedback current amplifier of Fig. 11.30(a) to meet the following specifications:

(i) $A_f \equiv I_o/I_s = -100$ A/A

(ii) amount of feedback = 60 dB

(iii) $R_{in} = 1\ \text{k}\Omega$

Specify the values of R_1, R_2, and μ. Assume that the amplifier μ has infinite input resistance and that $R_s = \infty$. For the MOSFET, $g_m = 5$ mA/V and $r_o = 20$ kΩ. What R_{out} is obtained?

11.73 Consider the feedback current amplifier in Fig. 11.30(a) (which was analyzed in Example 11.10). Let $R_s = R_{id} = \infty$. By setting $I_s = 0$ and breaking the feedback loop at the gate of Q, find an expression for the loop gain $A\beta$. Evaluate $A\beta$ for the component values given in Example 11.10 and hence determine A and A_f. Why do the results differ somewhat from those found in Example 11.10?

VE 11.4 **11.74** The feedback current amplifier in Fig. P11.74 utilizes two identical NMOS transistors sized so that at $I_D = 0.2$ mA they operate at $V_{OV} = 0.2$ V. Both devices have $V_t = 0.5$ V and $V_A = 10$ V.

(a) If I_s has zero dc component, show that both Q_1 and Q_2 are operating at $I_D = 0.2$ mA. What is the dc voltage at the input?
(b) Find the ideal value of $A_f = I_o/I_s$, and the value of β.
(c) Find g_m and r_o for each of Q_1 and Q_2.
(d) Find the A circuit and the value of R_i, A, and R_o.
(e) Find $A\beta$ and A_f.
(f) Find R_{in} and R_{out}.

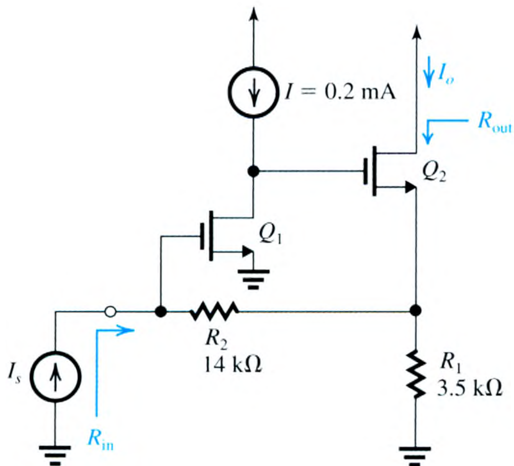


Figure P11.74

11.75 For the feedback current amplifier in Fig. P11.46:

(a) Find expressions for the ideal closed-loop gain I_o/I_s, and β. Also, give the β circuit.
(b) Provide the A circuit and derive expressions for R_i and A. Neglect r_o of both transistors.
(c) Find an expression for $A\beta$.
(d) For $g_{m1} = g_{m2} = 4$ mA/V, $R_D = 30$ kΩ, $R_M = 200$ Ω, and $R_F = 1$ kΩ, find the values of A, β, $A\beta$, A_f, $A_f|_{ideal}$, R_i, and R_{in}.
(e) If $r_{o2} = 20$ kΩ and $R_L = 1$ kΩ, find the output resistance as seen by R_L.

***11.76** The feedback current amplifier in Fig. P11.76(a) can be thought of as a "super" CG transistor. Note that rather than connecting the gate of Q_2 to signal ground, an amplifier is placed between source and gate. As you will soon find out, this negative feedback has the effect of increasing g_{m2} by a factor equal to μ.

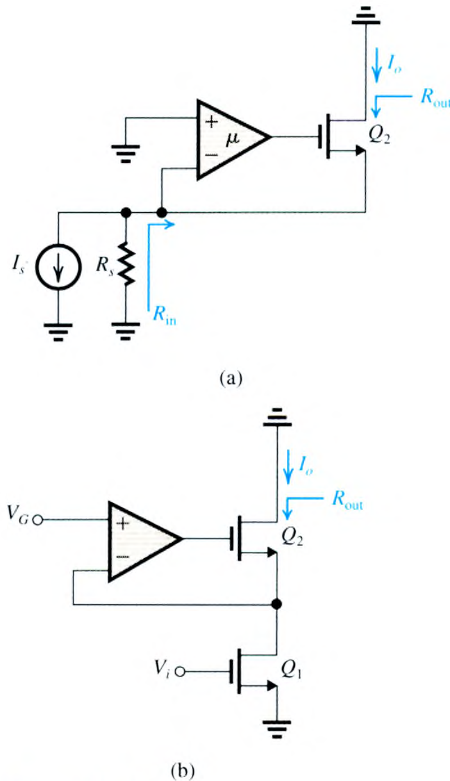


Figure P11.76

(a) If μ is infinite, what is the signal voltage at the input terminal? What is the input resistance? What is the current gain I_o/I_s?
(b) What is the value of β? Also, give the β circuit.
(c) For finite μ but assuming that the input resistance of the amplifier μ is very large, find the A circuit and derive expressions for A, R_i, and R_o.
(d) Find $A\beta$ and A_f. If μ is large, what is the value of A_f?
(e) Find R_{in} and R_{out}. Compare to those of a CG transistor.
(f) The "super" CG transistor can be utilized in the cascode configuration as shown in Fig. P11.76(b), where V_G is a dc bias voltage. Replacing Q_1 by its small-signal model, use the analogy of the resulting circuit to that in Fig. P11.76(a) to find I_o and R_{out}. Compare to the regular cascode.

***11.77** Figure P11.77 shows an interesting and very useful application of feedback to improve the performance of the current mirror formed by Q_1 and Q_2. Rather than connecting the drain of Q_1 to the gate, as is the case in simple current mirrors, an amplifier of gain $+\mu$ is connected between the drain and the gate. Note that the feedback loop does *not* include transistor Q_2. The feedback loop ensures that the value of the gate-to-source voltage of Q_1 is such that I_{o1} equals I_s. This regulated V_{gs} is also applied to Q_2. Thus, if W/L of Q_2 is n times W/L of Q_1, $I_{o2} = nI_{o1} = nI_s$. This current tracking, however, is *not* regulated by the feedback loop.

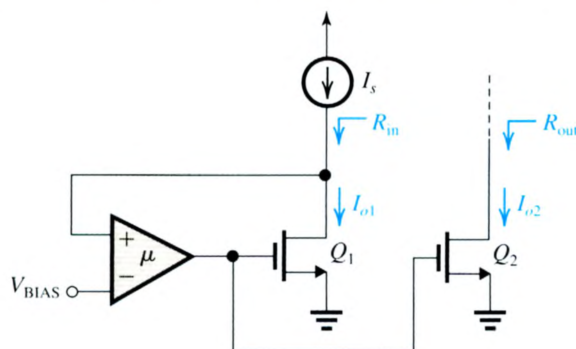


Figure P11.77

(a) Show that the feedback is negative.
(b) If μ is very large and the input resistance of the amplifier μ is infinite, what dc voltage appears at the drain of Q_1? If Q_1 is to operate at an overdrive voltage of 0.2 V, what is the minimum value that V_{BIAS} must have?
(c) Replacing Q_1 by its small-signal model, find an expression for the small-signal input resistance R_{in} assuming finite gain but infinite input resistance for the amplifier μ. Note that here it is much easier to do the analysis directly than to use the feedback-analysis approach. For large μ, what does R_{in} become?
(d) What is the output resistance R_{out}?

***11.78** The circuit in Fig. P11.78 is an implementation of a particular circuit building block known as **second-generation current convoyer** (CCII). It has three terminals besides ground: x, y, and z. The heart of the circuit is the feedback amplifier consisting of the differential amplifier μ and the complementary source follower (Q_N, Q_P). (Note that this feedback circuit is one we have encountered a number of times in this chapter, albeit with only one source-follower transistor.) In the following, assume that the differential amplifier has a very large gain μ (ideally infinite) and infinite differential input resistance. Also, let the two current mirrors have unity current-transfer ratios.

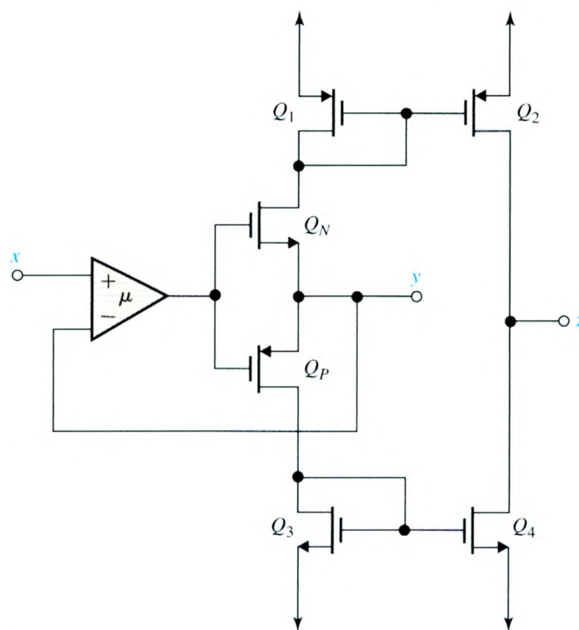


Figure P11.78

(a) If a resistance R is connected between y and ground, a voltage signal V_x is connected between x and ground, and z is short-circuited to ground. Find the current I_z through the short circuit. Show how this current is developed and its path for V_x positive and for V_x negative.
(b) If x is connected to ground, a current source I_y is connected to input terminal y, and z is connected to ground, what voltage appears at y and what is the input resistance seen by I_y? What is the current I_z that flows

through the output short circuit? Also, explain the current flow through the circuit for I_y positive and for I_y negative.

(c) What is the output resistance at z?

SIM *11.79 For the amplifier circuit in Fig. P11.79, assuming that V_s has a zero dc component, find the dc voltages at all nodes and the dc emitter currents of Q_1 and Q_2. Let the BJTs have $\beta = 100$. Use feedback analysis to find V_o/V_s and R_{in}. Let $V_{BE} = 0.7$ V.

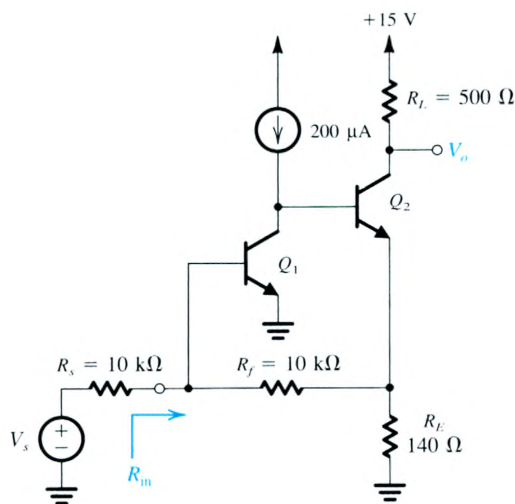


Figure P11.79

**11.80 Figure P11.80 shows a feedback amplifier utilizing the shunt–series topology. Here, the feedback network, formed by R_{E2} and R_{f1} senses the emitter current of Q_2, I_{e2}, which should be considered the output quantity of the feedback amplifier for the purpose of performing the feedback analysis. Of course, the output current I_{out} is proportional to I_{e2}. All transistors have $\beta = 100$ and $V_{BE} = 0.7$ V. Neglect r_o except in (g).

(a) Perform a dc analysis to find the dc emitter currents in Q_1 and Q_2 and hence determine their small-signal parameters.
(b) Replacing the BJTs with their hybrid-π models, give the equivalent circuit of the feedback amplifier.
(c) Find the ideal value of the closed-loop gain $A_f = I_{e2}/I_s$ where $I_s = V_s/R_s$. Hence, find the ideal value of the current gain I_{out}/I_{in}.
(d) Find β and give the β circuit.
(e) Give the A circuit and determine A, R_i, and R_o. Note that R_o is the resistance determined by breaking the emitter loop of Q_2 and measuring the resistance between the terminals thus created.
(f) Find $A\beta$, $1 + A\beta$, A_f, R_{if}, and R_{of}. Note that R_{of} represents the resistance that in effect appears in the emitter of Q_2 as a result of the feedback.
(g) Determine R_{in}, I_{out}/I_{in}, and R_{out}. To determine R_{out}, use $V_{A2} = 75$ V and recall that the maximum possible output resistance looking into the collector of a BJT

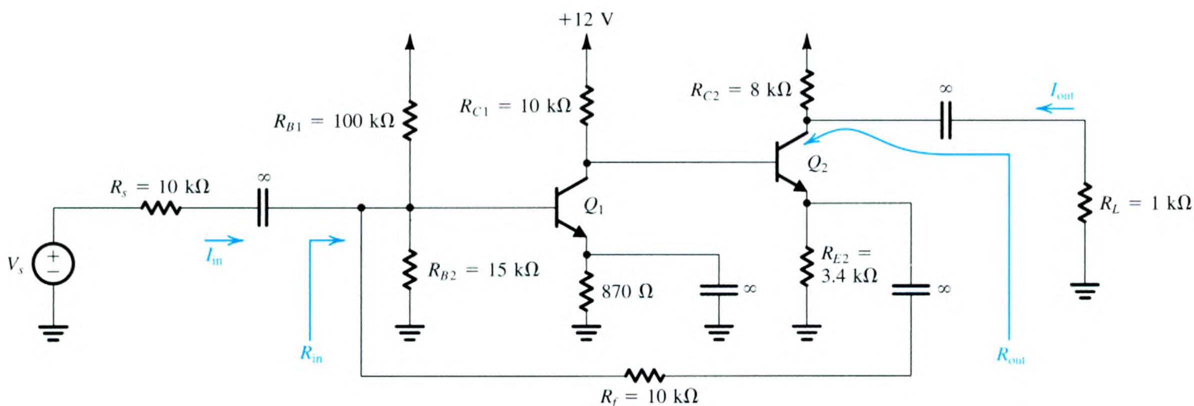


Figure P11.80

is approximately βr_o, where β is the BJT's β (see Problem 11.56).

Section 11.7: The Stability Problem

11.81 A feedback amplifier has an open-loop gain $A = 10^4$ V/V with a pole at 100 rad/s and two coincident poles at 10^4 rad/s. The feedback network is resistive with a feedback factor β. Find the frequency of 180° phase shift, the maximum value that β can have before oscillations start, and the corresponding minimum allowable closed-loop gain.

11.82 An op amp having a low-frequency gain of 10^5 V/V and a single-pole rolloff at 10^3 rad/s is connected in a negative-feedback loop via a feedback network having a transmission k and a two-pole rolloff at 10^3 rad/s. Find the value of k above which the closed-loop amplifier becomes unstable.

11.83 Consider a feedback amplifier for which the open-loop gain $A(s)$ is given by

$$A(s) = \frac{10,000}{(1 + s/10^4)(1 + s/10^5)^2}$$

If the feedback factor β is independent of frequency, find the frequency at which the phase shift is 180°, and find the critical value of β at which oscillation will commence.

Section 11.8: Effect of Feedback on the Amplifier Poles

11.84 A dc amplifier having a single-pole response with pole frequency 1 kHz and unity-gain frequency of 10 MHz is operated in a loop whose frequency-independent feedback factor is 0.1. Find the low-frequency gain, the 3-dB frequency, and the unity-gain frequency of the closed-loop amplifier. By what factor does the pole shift?

11.85 An amplifier has dc open-loop gain of 80 dB and a single pole with 100-Hz frequency. It is utilized to design a feedback amplifier with a 3-dB frequency of 10 kHz. What β is needed? What is the dc closed-loop gain realized? Give an expression for $A_f(s)$.

11.86 An amplifier having a low-frequency gain of 10^4 and poles at 10^3 Hz and 10^4 Hz is operated in a closed negative-feedback loop with a frequency-independent β.

(a) For what value of β do the closed-loop poles become coincident? At what frequency?

(b) What is the low-frequency, closed-loop gain corresponding to the situation in (a)? What is the value of the closed-loop gain at the frequency of the coincident poles?

(c) What is the value of Q corresponding to the situation in (a)?

(d) If β is increased by a factor of 10, what are the new pole locations? What is the corresponding pole Q?

D 11.87 A dc amplifier has an open-loop gain of 1000 and two poles, a dominant one at 10 kHz and a high-frequency one whose location can be controlled. It is required to connect this amplifier in a negative-feedback loop that provides a dc closed-loop gain of 20 and a maximally flat response. Find the required value of β and the frequency at which the second pole should be placed. What is the 3-dB frequency of the closed-loop amplifier?

D 11.88 A feedback amplifier having a dc closed-loop gain of 10 and a maximally flat second-order response with a 3-dB frequency of 1 kHz is required. The open-loop amplifier utilizes a cascade of two identical amplifier stages, each having a single-pole frequency response. Find the value required for β, and the 3-dB frequency and the dc gain of each of the two amplifier stages. Give an expression for $A_f(s)$. (*Hint*: For the maximally flat case, $\omega_{3dB} = \omega_0$.)

***11.89** Three identical inverting amplifier stages, each characterized by a low-frequency gain K and a single-pole response with $f_{3dB} = 100$ kHz, are connected in a feedback loop with $\beta = 1$. What is the minimum value of K at which the circuit oscillates? What would the frequency of oscillation be?

Section 11.9: Stability Study Using Bode Plots

11.90 Reconsider Exercise 11.21 for the case of the op amp wired as a unity-gain buffer. At what frequency is $|A\beta| = 1$? What is the corresponding phase margin?

11.91 Reconsider Exercise 11.21 for the case of a manufacturing error introducing a second pole at 10^3 Hz. What is now the frequency for which $|A\beta| = 1$? What is the corresponding phase margin? For what values of β is the phase margin 45° or more? What is the corresponding value of closed-loop gain?

11.92 (a) Use Eq. (11.48) to show that the peaking factor $P \equiv |A_f(j\omega_1)|/|A_f(0)|$ is related to the phase margin ϕ by

$$\phi = \cos^{-1}\left(1 - \frac{1}{2P^2}\right)$$

(b) Use the formula derived in (a) to find the phase margin that results in a peaking of (i) 5%, (ii) 10%, (iii) 0.1 dB, (iv) 1 dB, (v) 3 dB.

11.93 For the amplifier described by Fig. 11.38 and with frequency-independent feedback, what is the minimum closed-loop voltage gain that can be obtained for phase margins of 90° and 60°?

11.94 An amplifier has a dc gain of 10^4 and poles at 10^5 Hz, 3.16×10^5 Hz, and 10^6 Hz. Find the value of β, and the corresponding closed-loop gain, for which a phase margin of 45° is obtained.

***11.95** A two-pole amplifier for which $A_0 = 10^3$ and having poles at 1 MHz and 10 MHz is to be connected as a differentiator. On the basis of the rate-of-closure rule, what is the smallest differentiator time constant for which operation is stable?

Section 11.10: Frequency Compensation

D 11.96 A multipole amplifier having a first pole at 2 MHz and a dc open-loop gain of 80 dB is to be compensated for closed-loop gains as low as unity by the introduction of a new dominant pole. At what frequency must the new pole be placed?

D 11.97 For the amplifier described in Problem 11.96, rather than introducing a new dominant pole we can use additional capacitance at the circuit node at which the pole is formed to reduce the frequency of the first pole. If the frequency of the second pole is 20 MHz and if it remains unchanged while additional capacitance is introduced as mentioned, find the frequency to which the first pole must be lowered so that the resulting amplifier is stable for closed-loop gains as low as unity. By what factor is the capacitance at the controlling node increased?

11.98 For the amplifier whose $A(s)$ is depicted in Fig. 11.39, to what value must the first pole frequency be lowered to obtain stable performance for (a) $\beta = 0.001$ and (b) $\beta = 0.1$?

D11.99 An op amp with an open-loop voltage gain of 80 dB and poles at 10^5 Hz, 10^6 Hz, and 2×10^6 Hz is to be compensated to be stable for unity β. Assume that the op amp incorporates an amplifier equivalent to that in Fig. 11.41, with $C_1 = 150$ pF, $C_2 = 5$ pF, and $g_m = 40$ mA/V, and that f_{P1} is caused by the input circuit and f_{P2} by the output circuit of this amplifier. Find the required value of the compensating Miller capacitance and the new frequency of the output pole.

11.100 Contemplate the effects of pole splitting by considering Eqs. (11.54), (11.58), and (11.59) under the conditions that $R_1 \simeq R_2 = R$, $C_2 \simeq C_1/10 = C$, $C_f \gg C$, and $g_m = 100/R$, by calculating ω_{P1}, ω_{P2}, and ω'_{P1}, ω'_{P2}. Comment on the results.

11.101 An op amp has a dc gain of 120 dB and three poles: $\omega_{P1} = 10^5$ rad/s, $\omega_{P2} = 10^7$ rad/s, and $\omega_{P3} = 10^8$ rad/s. The frequency response must be compensated so that the gain has a uniform −20 dB/decade rolloff down to 0 dB and the 3-dB frequency as high as possible.

Miller compensation is used around one of the stages of the amplifier as shown in Fig. 11.41. Assume that without C_f connected, the input circuit is responsible for the first pole and the output circuit is responsible for the second pole. Let $C_1 = 1$ pF, $C_2 = 0.05$ pF, and $g_m = 10$ mA/V. When C_f is connected, assume that the second pole moves to a frequency higher than that of the third pole.

(a) Provide a neat and clearly labeled sketch of the Bode plot of the magnitude of the open-loop gain of the uncompensated amplifier. The frequency axis should extend from 10 rad/s to 10^9 rad/s.
(b) On the same figure, sketch the Bode plot for the magnitude of the gain of the compensated amplifier. Specify the frequency of the dominant pole and the unity-gain bandwidth of the compensated amplifier.
(c) Give the required value of C_f.
(d) What is the frequency to which the second pole moves?

CHAPTER 12

Output Stages and Power Amplifiers

IN THIS CHAPTER YOU WILL LEARN

- The classification of amplifier output stages based on the fraction of the cycle of an input sine wave during which the transistor conducts.
- Analysis and design of a variety of output-stage types ranging from the simple but power-inefficient emitter follower (class A) to the popular push–pull class AB circuit in both bipolar and CMOS technologies.
- Useful and interesting circuit techniques used in the design of power amplifiers.

Introduction

This chapter describes amplifier stages with large power gain. They must accommodate large voltage and/or current swings at their output. When used on their own, these stages are **power amplifiers**. When used as the final stage of a multistage amplifier, they are **output stages**. Generating these large swings, while at the same time maintaining linearity, is a fundamental challenge in the design of such stages. Audio amplifiers are a classic example: the voltage and current they deliver to a speaker must be sufficient to produce clearly audible sounds. However, human ears are capable of detecting even very subtle distortions in an audio waveform, so the power amplifier must also be highly linear. A high-quality audio amplifier can produce a sinusoidal output whose harmonic content is only a small fraction of a percent relative to the fundamental.

The most challenging requirement in the design of an output stage is that it deliver the required power in an *efficient* manner. Recall that an amplifier's efficiency is the fraction of power delivered to the amplifier (primarily from its power supplies), which is ultimately delivered to the load. To achieve high efficiency, the power dissipated in the output-stage transistors must be as low as possible. This prevents overheating the amplifier transistors, for which there is a maximum operating temperature (in the range of 150°C to 200°C for silicon devices). High efficiency can also prolong battery life; this is of paramount importance in, for example, the power amplifiers driving antennae and headphone speakers in mobile phones.

We begin this chapter with a study of the various output-stage configurations used in amplifiers that handle both low and high power. In this context, "high power" generally means greater than 1 W. Examples include the transmitter of a cell phone, which is typically required to deliver 1 W of power to its antenna, and a stereo system that delivers hundreds of watts of audio power to its speakers. Examples of "low-power" applications that may use similar output stages include a wireless Bluetooth transmitter and an audio amplifier to drive headphones, both of which must deliver around 10 mW or less.

12.1 Classification of Output Stages

Output stages are classified according to the collector-current waveform generated when we apply an input signal.[1] Figure 12.1 shows the classification for a sinusoidal input signal. The class A stage, whose associated waveform is shown in Fig. 12.1(a), is biased at a current I_C that is greater than the amplitude of the signal current, $\hat{I}_c$. Thus, the transistor in a class A stage conducts for the entire cycle of the input signal, and the conduction angle is 360°. In contrast, the class B stage, whose associated waveform is shown in Fig. 12.1(b), is biased at zero dc current. So a transistor in a class B stage conducts for only half the cycle of the input sine wave, giving a conduction angle of 180°. As we will see later, the negative halves of the sinusoid are supplied by another transistor that also operates in the class B mode and conducts during the alternate half-cycles.

An intermediate class between A and B, appropriately named class AB, involves biasing the transistor at a nonzero dc current much smaller than the peak current of the sine-wave

[1]We focus initially an BJT amplifiers, although the same classification is applied to CMOS amplifiers with drain current replacing the collector current.

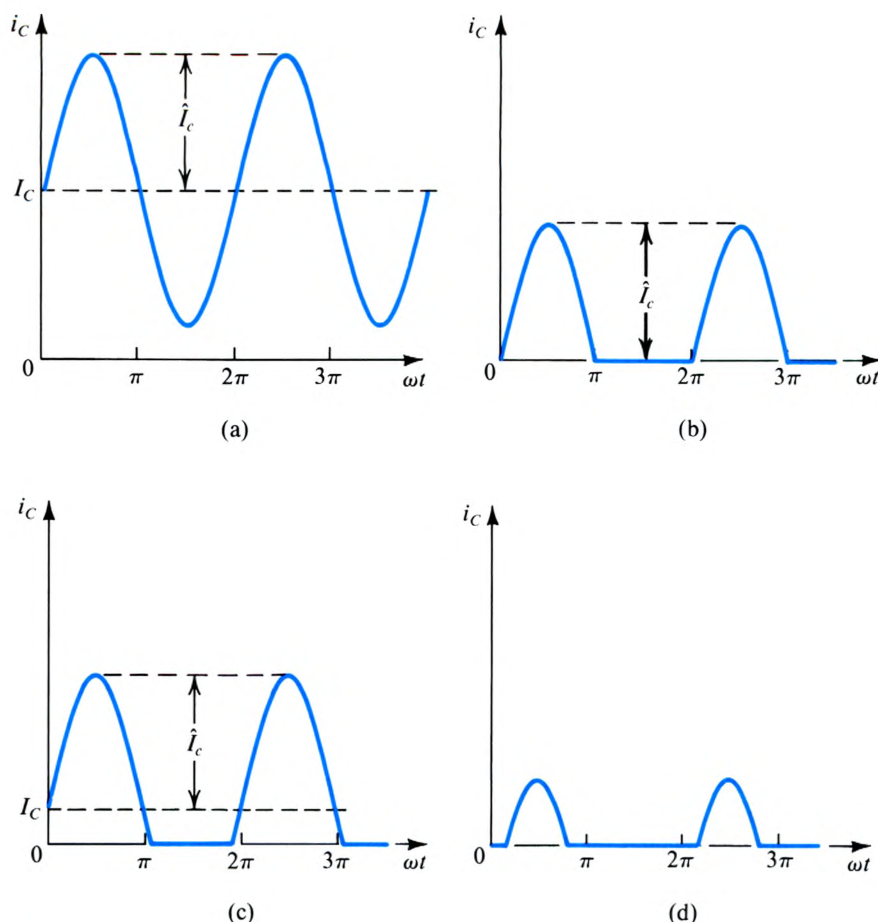


Figure 12.1 Collector-current waveforms for transistors operating in **(a)** class A, **(b)** class B, **(c)** class AB, and **(d)** class C amplifier stages.

signal. As a result, the transistor conducts for slightly more than half a cycle, as illustrated in Fig. 12.1(c). The resulting conduction angle is greater than 180° but much less than 360°. We then use another transistor that conducts for an interval slightly greater than that of the negative half-cycle, and the currents from the two transistors are combined in the load. During the intervals near the zero crossings of the input sinusoid, both transistors conduct.

A variation on the class AB, called the class G/H amplifier, utilizes two pairs of power supplies (e.g., the regular ±30-V supply and a higher-voltage supply of ±70 V). The higher-voltage supply is called upon only occasionally—for instance, to provide a short burst of high output power for a drum roll. Since the high-voltage supply operates infrequently, it can be of a low-cost design. Also, this is a more power-efficient arrangement than would be obtained if a class AB circuit were used and operated continuously from the higher-voltage supply.

Figure 12.1(d) shows the collector-current waveform for a transistor operated as a class C amplifier. The transistor conducts less than a half-cycle; that is, the conduction angle is less than 180°. The result is the periodically pulsating current waveform shown. To obtain a sinusoidal output voltage, this current is passed through a parallel *LC* circuit, tuned to the

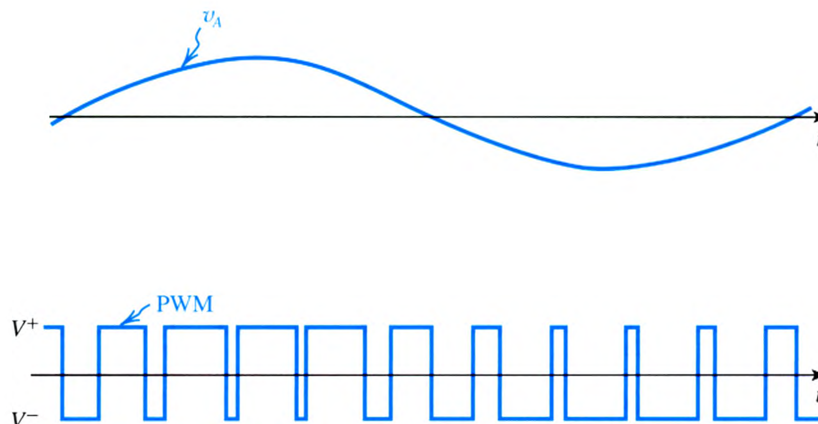


Figure 12.2 The PWM signal is made of rectangular pulses at constant frequency, whose width varies in proportion to an audio signal, V_A.

frequency of the input sinusoid. The tuned circuit acts as a bandpass filter (Chapter 14) and provides an output voltage proportional to the amplitude of the fundamental component in the Fourier-series representation of the current waveform. Class C amplifiers are intended for use where the signal is primarily sinusoidal, such as in radio-frequency transmitters. The tuned-resonator oscillator circuits in Chapter 15 also inherently operate in class C mode.

Class D stages take a very different approach to power amplification. Transistors are operated only as on-off switches at a frequency much higher than the signal to be amplified. The result is a square waveform whose average value varies in proportion to the signal being amplified. A popular method for doing this is pulsewidth modulation (PWM), illustrated in Fig. 12.2. Ideally, the transistors always exhibit either zero current (when they are off) or zero voltage drop (when they are on). Thus, they dissipate no power, and the class D audio amplifier can approach 100% efficiency. Naturally, real transistors have finite on-resistance. Moreover, during switching there can be a momentary short created between the supply voltage and ground. In addition, it takes some power to turn the output stage transistors on and off. Thus, class D stage efficiency is limited to values of 85–95%, still well above the efficiency of class A, B, and AB stages.

We study class A, B, and AB amplifiers in this chapter. They are used as output stages of op amps and audio power amplifiers. In the latter application, class AB is the preferred choice, for reasons we will explain in the sections that follow.

12.2 Class A Output Stage

The emitter follower is a popular class A stage. We have already seen, in Chapter 7, that it offers low small-signal output resistance. Here, we consider its large signal operation and see how its low output resistance affords good linearity.

12.2.1 Transfer Characteristic

Figure 12.3 shows an emitter follower Q_1 biased with a constant current I supplied by transistor Q_2. Since the emitter current $i_{E1} = I + i_L$, the bias current I must be greater than

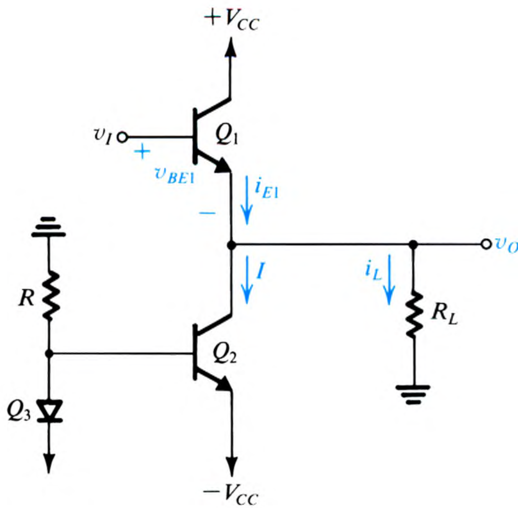


Figure 12.3 An emitter follower (Q_1) biased with a constant current I supplied by transistor Q_2.

the largest negative load current; otherwise, Q_1 cuts off and class A operation is no longer maintained.

The transfer characteristic of the emitter follower of Fig. 12.3 is described by

$$v_O = v_I - v_{BE1} \tag{12.1}$$

where v_{BE1} depends on the emitter current i_{E1} and thus on the load current i_L. If we neglect the relatively small changes in v_{BE1} (60 mV for every factor-of-10 change in emitter current), we get the linear transfer curve in Fig. 12.4. As indicated, the positive limit of the linear region is determined by the saturation of Q_1; thus

$$v_{O\max} = V_{CC} - V_{CE1\text{sat}} \tag{12.2}$$

This requires v_I to exceed V_{CC}. Therefore, $v_{O\max}$ may be reduced if v_I is limited by the preceding stage.

In the negative direction, depending on the values of I and R_L, the limit of the linear region is determined either by Q_1 turning off,

$$v_{O\min} = -IR_L \tag{12.3}$$

or by Q_2 saturating,

$$v_{O\min} = -V_{CC} + V_{CE2\text{sat}} \tag{12.4}$$

Of these, the limit in (12.4) is lower and therefore provides a wider linear range. It is achieved as long as the bias current I is greater than the load current corresponding to Eq. (12.4),

$$I \geq \frac{|-V_{CC} + V_{CE2\text{sat}}|}{R_L} \tag{12.5}$$

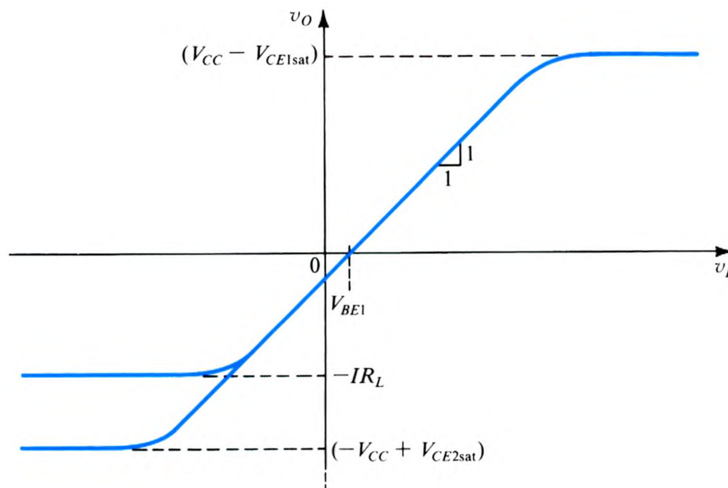


Figure 12.4 Transfer characteristic of the emitter follower in Fig. 12.3. We obtain a linear characteristic by neglecting the change in v_{BE1} with i_L. The maximum positive output is determined by the saturation of Q_1. In the negative direction, the limit of the linear region is determined either by Q_1 turning off or by Q_2 saturating, depending on the values of I and R_L.

Example 12.1

Any change in the slope of the emitter follower's transfer characteristic, Fig. 12.4, results in distortion. Here, we perform small-signal analysis to find the incremental gain, hence slope, at several operating points. In Fig. 12.3, let $V_{CC} = 15$ V, $I = 120$ mA, and $R_L = 100\ \Omega$. Find the incremental (small-signal) gain of the circuit at $v_O = +10$ V, 0 V, and -10 V.

Solution

We saw in Chapter 7 that (assuming $r_{o1} \gg R_L$) the small-signal voltage gain of the emitter follower is

$$A_v = \frac{R_L}{R_L + r_{e1}} = \frac{1}{1 + r_{e1}/R_L}$$

As long Q_1 remains in active mode, we may use this result to find the incremental gain. The gain remains very close to unity as long as r_{e1} remains much smaller than R_L. However, r_{e1} depends on the output voltage v_O,

$$r_{e1} = \frac{V_T}{i_{E1}} = \frac{V_T}{I + i_L} = \frac{V_T}{I + v_O/R_L}$$

Assuming $V_T = 25$ mV, we use these equations to find the small-signal gain at $v_O = +10$ V, 0 V, and -10 V as follows:

v_O	$i_L = v_O/R_L$	$i_{E1} = I + i_L$	$r_{e1} = V_T/i_{E1}$	A_v
+10 V	100 mA	220 mA	0.11 Ω	0.999 V/V
0 V	0 mA	120 mA	0.2 Ω	0.998 V/V
−10 V	−100 mA	20 mA	1.25 Ω	0.988 V/V

In all three cases, $i_{E1} > 0$, and the Q_1 collector voltage, $+V_{CC}$, exceeds the emitter voltage, v_O, by at least 5 V, so Q_1 does indeed remain in active mode. The incremental gain varies by only about 1%.

Example 12.1 shows that at low output voltages, i_{E1} decreases and r_{e1} increases, affecting the slope, A_v. In the example, r_{e1} remains much less than R_L over the range $|v_O| < 10$ V, so the linearity is relatively good. If we want to further improve linearity, we must increase the bias current I. Unfortunately, we will see that this also increases power dissipation in Q_1 and Q_2 and lowers efficiency.

EXERCISES

D12.1 For the emitter follower in Fig. 12.3, $V_{CC} = 15$ V, $V_{CEsat} = 0.2$ V, $V_{BE} = 0.7$ V and constant, and β is very high. Find the value of R that will establish a bias current large enough to allow the largest possible output signal swing for $R_L = 1$ kΩ. Determine the resulting output signal swing and the minimum and maximum emitter currents for Q_1.
Ans. 0.97 kΩ; –14.8 V to +14.8 V; 0 to 29.6 mA

12.2 For the emitter follower of Exercise 12.1, in which $I = 14.8$ mA and $R_L = 1$ kΩ, consider the case in which v_O is limited to the range –10 V to +10 V. Let Q_1 have $v_{BE} = 0.6$ V at $i_C = 1$ mA, and assume $\alpha \simeq 1$. Find v_I corresponding to $v_O = -10$ V, 0 V, and +10 V. At each of these points, use small-signal analysis to find the voltage gain v_o/v_i. Note that the incremental voltage gain gives the slope of the v_O-versus-v_I characteristic.
Ans. –9.36 V, 0.67 V, 10.68 V; 0.995 V/V, 0.998 V/V, 0.999 V/V

12.2.2 Signal Waveforms

Consider the operation of the emitter-follower circuit of Fig. 12.3 for sine-wave input. Neglecting V_{CEsat}, we see that if the bias current I is properly selected, the output voltage can swing from $-V_{CC}$ to $+V_{CC}$ with the quiescent value being zero, as shown in Fig. 12.5(a). Figure 12.5(b) shows the corresponding waveform of $v_{CE1} = V_{CC} - v_O$. Now, assuming that the bias current I is selected to allow a maximum negative load current of V_{CC}/R_L, that is,

$$I = V_{CC}/R_L$$

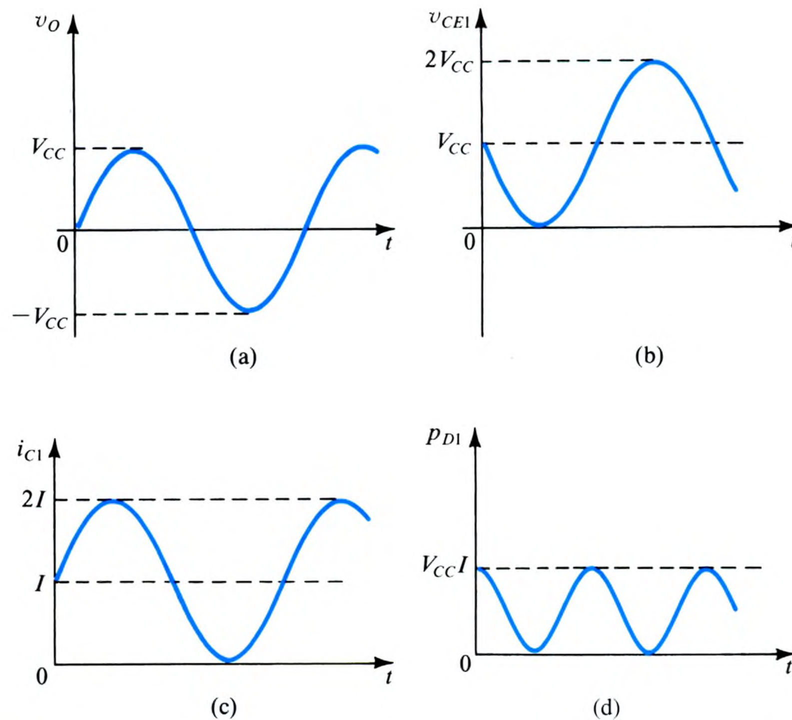


Figure 12.5 Maximum signal waveforms in the class A output stage of Fig. 12.3 under the condition $I = V_{CC}/R_L$ or, equivalently, $R_L = V_{CC}/I$. Note that we have neglected the transistor saturation voltages.

the collector current of Q_1 will have the waveform in Fig. 12.5(c). Finally, Fig. 12.5(d) shows the waveform of the **instantaneous power dissipation** in Q_1,

$$p_{D1} \equiv v_{CE1} i_{C1} \tag{12.6}$$

12.2.3 Power Dissipation

Figure 12.5(d) shows that the maximum instantaneous power dissipation in Q_1 is $V_{CC}I$. This is equal to the power dissipation in Q_1 with no input signal applied, that is, the quiescent power dissipation. Thus the emitter-follower transistor dissipates the largest amount of power when $v_O = 0$. Since we could easily have no input signal for prolonged periods of time, transistor Q_1 must be able to withstand a continuous power dissipation of $V_{CC}I$.

The power dissipation in Q_1 depends on the value of R_L. Consider the extreme case of an output open circuit, that is, $R_L = \infty$. In this case, $i_{C1} = I$ is constant and the instantaneous power dissipation in Q_1 will depend on the instantaneous value of v_O. The maximum power dissipation will occur when $v_O = -V_{CC}$, for in this case v_{CE1} is a maximum of $2V_{CC}$ and $p_{D1} = 2V_{CC}I$. This condition, however, would not normally persist for a prolonged interval, so the design need not be that conservative. Observe that with an open-circuit load, the average power dissipation in Q_1 is $V_{CC}I$. A far more dangerous situation occurs at the other extreme of R_L—specifically, $R_L = 0$. In the event of an output short circuit, a positive input voltage would theoretically result in an infinite load current. In practice, a very large current may flow through Q_1, and if the short-circuit condition persists, the resulting large power dissipation

in Q_1 can raise its junction temperature beyond the maximum allowed, causing permanent damage. To guard against this, we usually equip output stages with **short-circuit protection**.

In designing an emitter-follower stage, we must also consider the power dissipation in Q_2. Since Q_2 conducts a constant current I, and the maximum value of v_{CE2} is $2V_{CC}$, the maximum instantaneous power dissipation in Q_2 is $2V_{CC}I$. This maximum, however, occurs when $v_O = V_{CC}$, a condition that would not normally prevail for a prolonged period of time. A more significant quantity for design purposes is the average power dissipation in Q_2, which is $V_{CC}I$.

Example 12.2

Consider the emitter follower in Fig. 12.3 with $V_{CC} = 10$ V, $I = 100$ mA, and $R_L = 100\,\Omega$.

(a) Find the power dissipated in Q_1 and Q_2 under quiescent conditions $(v_O = 0)$.

(b) For a sinusoidal output voltage of maximum possible amplitude (neglecting V_{CEsat}), find the average power dissipation in Q_1 and Q_2. Also find the load power.

Solution

(a) Under quiescent conditions $v_O = 0$, and each of Q_1 and Q_2 conducts a current $I = 100$ mA $= 0.1$ A and has a voltage $V_{CE} = V_{CC} = 10$ V, thus

$$P_{D1} = P_{D2} = V_{CC}I = 10 \times 0.1 = 1 \text{ W}$$

(b) For a sinusoidal output voltage of maximum possible amplitude (i.e., 10-V peak), the instantaneous power dissipation in Q_1 will be as shown in Fig. 12.5(d). Thus the average power dissipation in Q_1 will be

$$P_{D1} = \frac{1}{2}V_{CC}I = \frac{1}{2} \times 10 \times 0.1 = 0.5 \text{ W}$$

For Q_2, the current is constant at $I = 0.1$ A and the voltage at the collector will have an average value of 0 V. Thus the average voltage across Q_2 will be V_{CC} and the average dissipation will be

$$\begin{aligned} P_{D2} &= I \times v_{CE}\big|_{\text{average}} \\ &= I \times V_{CC} = 0.1 \times 10 = 1 \text{ W} \end{aligned}$$

Finally, we can find the power delivered to the load from

$$\begin{aligned} P_L &= \frac{V_{orms}^2}{R_L} \\ &= \frac{(10/\sqrt{2})^2}{100} = 0.5 \text{ W} \end{aligned}$$

12.2.4 Power-Conversion Efficiency

The power-conversion efficiency of an output stage is defined as

$$\eta \equiv \frac{\text{Load power}(P_L)}{\text{Supply power}(P_S)} \tag{12.7}$$

For the emitter follower of Fig. 12.3, assuming that the output voltage is a sinusoid with the peak value $\hat{V}_o$, the average load power will be

$$P_L = \frac{(\hat{V}_o/\sqrt{2})^2}{R_L} = \frac{1}{2}\frac{\hat{V}_o^2}{R_L} \tag{12.8}$$

Since the current in Q_2 is constant (I), the power drawn from the negative supply[2] is $V_{CC}I$. The *average* current in Q_1 is equal to I, and thus the average power drawn from the positive supply is $V_{CC}I$. Thus the total average supply power is

$$P_S = 2V_{CC}I \tag{12.9}$$

Equations (12.8) and (12.9) can be combined to yield

$$\eta = \frac{1}{4}\frac{\hat{V}_o^2}{IR_L V_{CC}} = \frac{1}{4}\left(\frac{\hat{V}_o}{IR_L}\right)\left(\frac{\hat{V}_o}{V_{CC}}\right) \tag{12.10}$$

Since $\hat{V}_o \leq V_{CC}$ and $\hat{V}_o \leq IR_L$, we reach maximum efficiency when

$$\hat{V}_o = V_{CC} = IR_L \tag{12.11}$$

The maximum efficiency attainable is 25%. Because this is rather low, we rarely use the class A output stage in high-power applications (>1 W). Note also that in practice the output voltage swing is limited to lower values to avoid transistor saturation and nonlinear distortion, as illustrated in Example 12.1. Thus the efficiency in practice is usually in the 10% to 20% range.

EXERCISE

12.3 For the emitter follower of Fig. 12.3, let $V_{CC} = 10$ V, I = 100 mA, and $R_L = 100\ \Omega$. If the output voltage is an 8-V-peak sinusoid, find: (a) the power delivered to the load; (b) the average power drawn from the supplies; (c) the power-conversion efficiency. Ignore the loss in Q_3 and R.
Ans. 0.32 W; 2 W; 16%

[2]This does *not* include the power drawn by the biasing resistor R and the diode-connected transistor Q_3.

12.3 Class B Output Stage

Figure 12.6 shows a class B output stage. It consists of a complementary pair of transistors (an *npn* and a *pnp*) connected so that both cannot conduct simultaneously.

12.3.1 Circuit Operation

When the input voltage v_I is zero, both transistors are cut off and the output voltage v_O is zero. As v_I goes positive and exceeds about 0.5 V, Q_N conducts and operates as an emitter follower. In this case v_O follows v_I (i.e., $v_O = v_I - v_{BEN}$) and Q_N supplies the load current. Meanwhile, the emitter–base junction of Q_P will be reverse biased by the V_{BE} of Q_N, which is approximately 0.7 V. Thus Q_P will be cut off.

If the input goes negative by more than about 0.5 V, Q_P turns on and acts as an emitter follower. Again v_O follows v_I (i.e., $v_O = v_I + v_{EBP}$), but in this case Q_P supplies the load current (in the direction opposite to that of i_L, since v_O will be negative), and Q_N will be cut off.

We conclude that the transistors in the class B stage of Fig. 12.6 are biased at zero current and conduct only when there is an input signal. The circuit operates in a **push–pull** fashion: Q_N *pushes* (sources) current into the load when v_I is positive, and Q_P *pulls* (sinks) current from the load when v_I is negative.

12.3.2 Transfer Characteristic

A sketch of the transfer characteristic of the class B stage is shown in Fig. 12.7. Note that there is a range of v_I centered around zero where both transistors are cut off and v_O is zero. This **dead band** results in the **crossover distortion** illustrated in Fig. 12.8 for the case of an input sine wave. The effect of crossover distortion is most pronounced when the amplitude of the input signal is small. Crossover distortion in audio power amplifiers gives rise to unpleasant sounds.

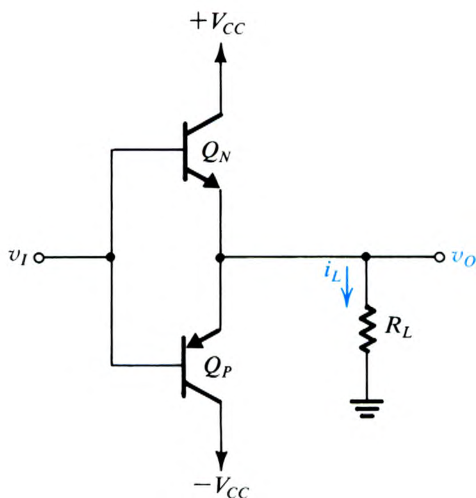


Figure 12.6 A class B output stage.

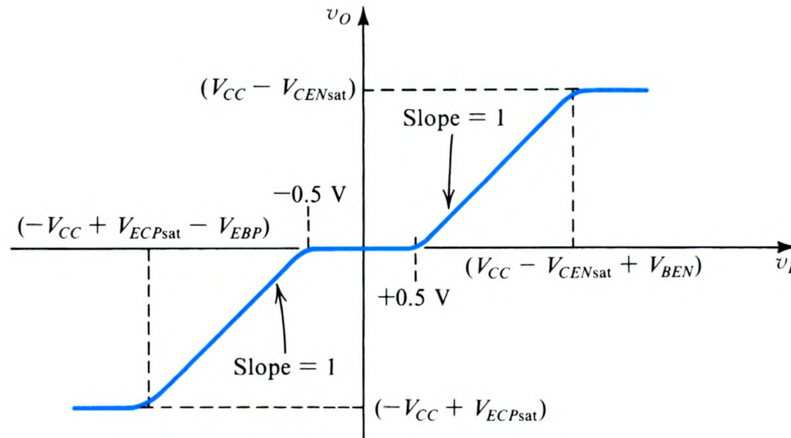


Figure 12.7 Transfer characteristic for the class B output stage in Fig. 12.6.

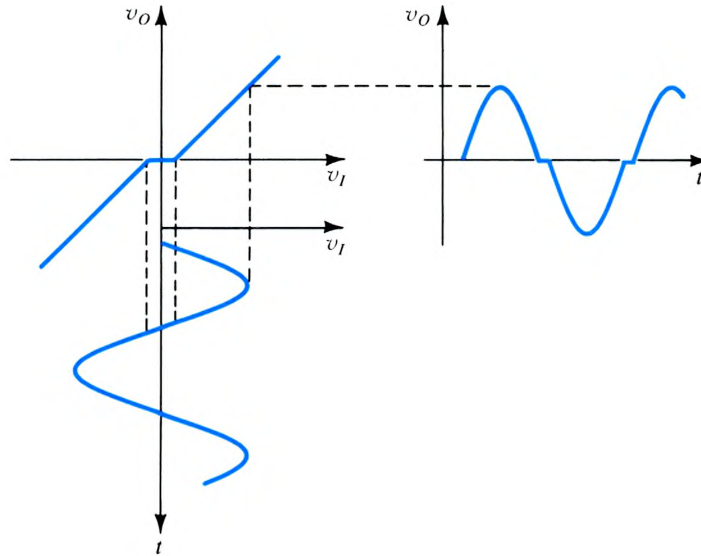


Figure 12.8 Illustrating how the dead band in the class B transfer characteristic results in crossover distortion.

12.3.3 Power-Conversion Efficiency

To calculate the power-conversion efficiency, η, of the class B stage, we neglect the crossover distortion and consider the case of an output sinusoid of peak amplitude $\hat{V}_o$. The average load power will be

$$P_L = \frac{1}{2}\frac{\hat{V}_o^2}{R_L} \tag{12.12}$$

The current drawn from each supply will consist of half-sine waves of peak amplitude $(\hat{V}_o/R_L)$. Thus the average current drawn from each of the two power supplies will be $\hat{V}_o/\pi R_L$.

This means that the average power drawn from each of the two power supplies will be the same,

$$P_{S+} = P_{S-} = \frac{1}{\pi}\frac{\hat{V}_o}{R_L}V_{CC} \tag{12.13}$$

and the total supply power will be

$$P_S = \frac{2}{\pi}\frac{\hat{V}_o}{R_L}V_{CC} \tag{12.14}$$

Thus the efficiency will be given by

$$\eta = \left(\frac{1}{2}\frac{\hat{V}_o^2}{R_L}\right) \Big/ \left(\frac{2}{\pi}\frac{\hat{V}_o}{R_L}V_{CC}\right) = \frac{\pi}{4}\frac{\hat{V}_o}{V_{CC}} \tag{12.15}$$

It follows that we reach the maximum efficiency when $\hat{V}_o$ is at its maximum. This maximum is limited by the saturation of Q_N and Q_P to $V_{CC} - V_{CEsat} \simeq V_{CC}$. At this value of peak output voltage, the power-conversion efficiency is

$$\eta_{\max} = \frac{\pi}{4} = 78.5\% \tag{12.16}$$

This value is much larger than in the class A stage (25%). Finally, note that we can find the maximum average power available from a class B output stage by substituting $\hat{V}_o = V_{CC}$ in Eq. (12.12),

$$P_{L\max} = \frac{1}{2}\frac{V_{CC}^2}{R_L} \tag{12.17}$$

12.3.4 Power Dissipation

Unlike the class A stage, which dissipates maximum power under quiescent conditions ($v_O = 0$), in the class B stage the quiescent power dissipation is zero. When an input signal is applied, the *average* power dissipated in the class B stage is given by

$$P_D = P_S - P_L \tag{12.18}$$

Substituting for P_S from Eq. (12.14) and for P_L from Eq. (12.12) gives

$$P_D = \frac{2}{\pi}\frac{\hat{V}_o}{R_L}V_{CC} - \frac{1}{2}\frac{\hat{V}_o^2}{R_L} \tag{12.19}$$

From symmetry we see that half of P_D is dissipated in Q_N and the other half in Q_P. Thus Q_N and Q_P must be capable of safely dissipating $\frac{1}{2}P_D$ watts. Since P_D depends on $\hat{V}_o$, we must find the worst-case power dissipation, $P_{D\max}$. Differentiating Eq. (12.19) with respect to $\hat{V}_o$ and setting the derivative to zero gives the value of $\hat{V}_o$ that results in maximum average power dissipation as

$$\hat{V}_o|_{P_{D\max}} = \frac{2}{\pi}V_{CC} \tag{12.20}$$

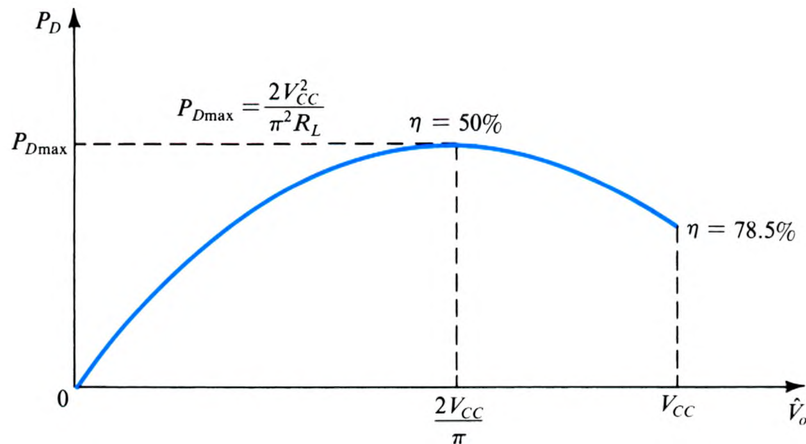


Figure 12.9 Power dissipation of the class B output stage versus amplitude of the output sinusoid.

Substituting this value in Eq. (12.19) gives

$$P_{Dmax} = \frac{2V_{CC}^2}{\pi^2 R_L} \tag{12.21}$$

Thus,

$$P_{DNmax} = P_{DPmax} = \frac{V_{CC}^2}{\pi^2 R_L} \tag{12.22}$$

At the point of maximum power dissipation, the efficiency can be evaluated by substituting for $\hat{V}_o$ from Eq. (12.20) into Eq. (12.15); hence, $\eta = 50\%$.

Figure 12.9 shows a sketch of P_D (Eq. 12.19) versus the peak output voltage $\hat{V}_o$. The data sheets of IC power amplifiers usually give curves like this, except that P_D is plotted versus $P_L = \frac{1}{2}(\hat{V}_o^2/R_L)$ rather than $\hat{V}_o$. Notice something interesting in Fig. 12.9: Increasing $\hat{V}_o$ beyond $2V_{CC}/\pi$ *decreases* the power dissipated in the class B stage while increasing the load power. The price we pay is an increase in nonlinear distortion as a result of approaching the saturation region of operation of Q_N and Q_P. Transistor saturation flattens the peaks of the output sine waveform. Unfortunately, this type of distortion cannot be significantly reduced by applying negative feedback (see Section 11.2), and thus transistor saturation should be avoided in applications requiring low distortion.

Example 12.3

We need to design a class B output stage to deliver an average power of 20 W to an 8-Ω load. We decide to select a power supply V_{CC} about 5 V greater than the peak output voltage. This avoids transistor saturation and the associated nonlinear distortion. Determine the supply voltage required, the peak current drawn from each supply, the total supply power, and the power-conversion efficiency. Also determine the maximum power that each transistor must be able to dissipate safely.

Solution

Since

$$P_L = \frac{1}{2}\frac{\hat{V}_o^2}{R_L}$$

then

$$\hat{V}_o = \sqrt{2P_LR_L} = \sqrt{2 \times 20 \times 8} = 17.9\text{ V}$$

Therefore we select $V_{CC} = 23$ V.

The peak current drawn from each supply is

$$\hat{I}_o = \frac{\hat{V}_o}{R_L} = \frac{17.9}{8} = 2.24\text{ A}$$

Since each supply provides a current waveform of half-sinusoids, the average current drawn from each supply will be $\hat{I}_o/\pi$. Thus the average power drawn from each supply is

$$P_{S+} = P_{S-} = \frac{1}{\pi} \times 2.24 \times 23 = 16.4\text{ W}$$

for a total supply power of 32.8 W. The power-conversion efficiency is

$$\eta = \frac{P_L}{P_S} = \frac{20}{32.8} \times 100 = 61\%$$

The maximum power dissipated in each transistor is given by Eq. (12.22); thus,

$$P_{DN\max} = P_{DP\max} = \frac{V_{CC}^2}{\pi^2 R_L}$$

$$= \frac{(23)^2}{\pi^2 \times 8} = 6.7\text{ W}$$

This occurs when the voltage swing is given by Eq. (12.20).

EXERCISE

12.4 For the class B output stage of Fig. 12.6, let $V_{CC} = 6$ V and $R_L = 4\ \Omega$. If the output is a sinusoid with 4.5-V peak amplitude, find (a) the output power; (b) the average power drawn from each supply; (c) the power efficiency obtained at this output voltage; (d) the peak currents supplied by v_I, assuming that $\beta_N = \beta_P = 50$; and (e) the maximum power that each transistor must be capable of dissipating safely.

Ans. (a) 2.53 W; (b) 2.15 W; (c) 59%; (d) 22.1 mA; (e) 0.91 W

12.4 Class AB Output Stage

So far, we have seen that a class A stage avoids distortion by ensuring some minimal collector current flows in the signal-path transistor at all times, whereas crossover distortion arises in a class B stage when the collector currents become zero. We can virtually eliminate crossover distortion by biasing the complementary output transistors of the class B output stage at a small nonzero current. The result is the class AB output stage in Fig. 12.10. We apply a bias voltage V_{BB} between the bases of Q_N and Q_P. For $v_I = 0$, $v_O = 0$, and a voltage $V_{BB}/2$ appears across the base–emitter junction of each of Q_N and Q_P. Assuming matched devices,

$$i_N = i_P = I_Q = I_S e^{V_{BB}/2V_T} \tag{12.23}$$

We select the value of V_{BB} to yield the required quiescent current I_Q.

12.4.1 Circuit Operation

When v_I goes positive by a certain amount, the voltage at the base of Q_N increases by the same amount, and the output becomes positive at an almost equal value,

$$v_O = v_I + \frac{V_{BB}}{2} - v_{BEN} \tag{12.24}$$

The positive v_O causes a current i_L to flow through R_L, and thus i_N must increase; that is,

$$i_N = i_P + i_L \tag{12.25}$$

The increase in i_N will be accompanied by a corresponding increase in v_{BEN} (above the quiescent value of $V_{BB}/2$). However, since the voltage between the two bases remains constant

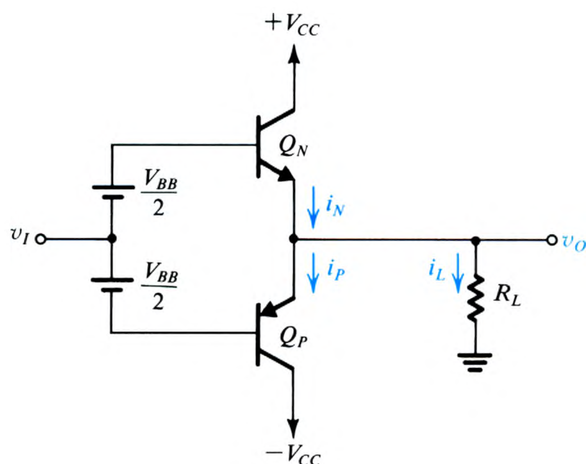


Figure 12.10 Class AB output stage. A bias voltage V_{BB} is applied between the bases of Q_N and Q_P, giving rise to a bias current I_Q given by Eq. (12.23). Thus, for small v_I, both transistors conduct and crossover distortion is almost completely eliminated.

at V_{BB}, the increase in v_{BEN} will result in an equal decrease in v_{EBP} and hence in i_P. We can derive the relationship between i_N and i_P as follows:

$$v_{BEN} + v_{EBP} = V_{BB}$$

$$V_T \ln\frac{i_N}{I_S} + V_T \ln\frac{i_P}{I_S} = 2V_T \ln\frac{I_Q}{I_S}$$

$$i_N i_P = I_Q^2 \tag{12.26}$$

Thus, as i_N increases, i_P decreases by the same ratio while the product remains constant. We can combine Eqs. (12.25) and (12.26) to yield i_N for a given i_L as the solution to the quadratic equation

$$i_N^2 - i_L i_N - I_Q^2 = 0 \tag{12.27}$$

From the equations above, we can see that for positive output voltages, the load current is supplied by Q_N, which acts as the output emitter follower. Meanwhile, Q_P will be conducting a current that decreases as v_O increases; for large v_O the current in Q_P can be ignored altogether.

For negative input voltages the opposite occurs: The load current will be supplied by Q_P, which acts as the output emitter follower, while Q_N conducts a current that gets smaller as v_I becomes more negative. Equation (12.26), relating i_N and i_P, holds for negative inputs as well.

We conclude that the class AB stage operates in much the same manner as the class B circuit, with one important exception: For small v_I, both transistors conduct, and as v_I is increased or decreased, one of the two transistors takes over the operation. Since the transition is a smooth one, crossover distortion is almost totally eliminated. Figure 12.11 shows the transfer characteristic of the class AB stage.

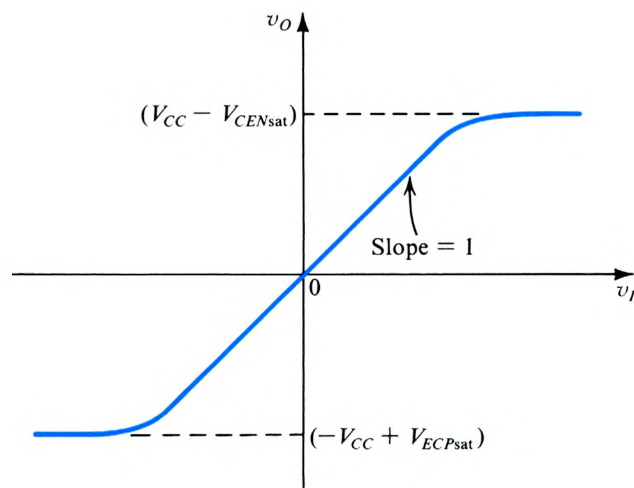


Figure 12.11 Transfer characteristic of the class AB stage in Fig. 12.10.

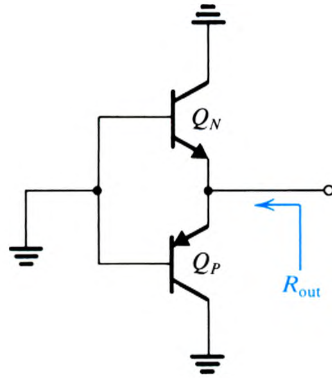


Figure 12.12 Determining the small-signal output resistance of the class AB circuit of Fig. 12.10.

The power relationships in the class AB stage are almost identical to those we derived for the class B circuit in Section 12.4. The only difference is that under quiescent conditions the class AB circuit dissipates a power of $V_{CC}I_Q$ per transistor. Since I_Q is usually much smaller than the peak load current, the quiescent power dissipation is usually small. Nevertheless, it can be taken into account easily. Specifically, we can simply add the quiescent dissipation per transistor to its maximum power dissipation with an input signal applied, to obtain the total power dissipation that the transistor must be able to handle safely.

12.4.2 Output Resistance

To limit distortion, we must ensure that the class AB stage small-signal output resistance remains small compared to the load resistance. If we assume that the source supplying v_I is ideal, then the output resistance of the class AB stage can be determined from the circuit in Fig. 12.12 as

$$R_{out} = r_{eN} \,\|\, r_{eP} \tag{12.28}$$

where r_{eN} and r_{eP} are the small-signal emitter resistances of Q_N and Q_P, respectively. At a given input voltage, we can find the currents i_N and i_P, and r_{eN} and r_{eP} are given by

$$r_{eN} = \frac{V_T}{i_N} \tag{12.29}$$

$$r_{eP} = \frac{V_T}{i_P} \tag{12.30}$$

Thus,

$$R_{out} = \frac{V_T}{i_N} \,\Big\|\, \frac{V_T}{i_P} = \frac{V_T}{i_P + i_N} \tag{12.31}$$

Since as i_N increases, i_P decreases, and vice versa, the output resistance remains approximately constant in the region around $v_I = 0$. This, in turn, keeps the incremental gain near unity and is the reason for the virtual absence of crossover distortion. At larger load currents, either i_N or i_P will be significant, and R_{out} decreases as the load current increases.

Example 12.4

In this example we explore the transfer characteristic, v_O versus v_I, of the class AB circuit in Fig. 12.10. For this purpose let $V_{CC} = 15$ V, $I_Q = 2$ mA, and $R_L = 100\ \Omega$. Assume that Q_N and Q_P are matched and have $I_S = 10^{-13}$ A. First, determine the required value of the bias voltage V_{BB}. Then, find the transfer characteristic for v_O in the range -10 V to $+10$ V.

Solution

To determine the required value of V_{BB} we use Eq. (12.23) with $I_Q = 2$ mA and $I_S = 10^{-13}$ A. Thus,

$$V_{BB} = 2V_T \ln(I_Q/I_S) = 2 \times 0.025 \ln\left(2 \times 10^{-3}/10^{-13}\right) = 1.186 \text{ V}$$

The easiest way to determine the transfer characteristic is to work backward; that is, for a given v_O we determine the corresponding value of v_I. We shall outline the process for positive v_O:

1. Assume a value for v_O.
2. Determine the load current i_L,

$$i_L = v_O/R_L$$

3. Use Eq. (12.27) to determine the current conducted by Q_N, i_N.
4. Determine v_{BEN} from

$$v_{BEN} = V_T \ln(i_N/I_S)$$

5. Determine v_I from

$$v_I = v_O + v_{BEN} - V_{BB}/2$$

It is also useful to find i_P and v_{EBP} as follows:

$$i_P = i_N - i_L$$

$$v_{EBP} = V_T \ln(i_P/I_S)$$

A similar process can be used for negative v_O. In fact, we can use symmetry and avoid repeating the calculations. The results are displayed in the following table:

Example 12.4 *continued*

v_O (V)	i_L (mA)	i_N (mA)	i_P (mA)	v_{BEN} (V)	v_{EBP} (V)	v_I (V)	v_O/v_I	R_{out} (Ω)	v_o/v_i
+10.0	100	100.04	0.04	0.691	0.495	10.1	0.99	0.25	1.00
+5.0	50	50.08	0.08	0.673	0.513	5.08	0.98	0.50	1.00
+1.0	10	10.39	0.39	0.634	0.552	1.041	0.96	2.32	0.98
+0.5	5	5.70	0.70	0.619	0.567	0.526	0.95	4.03	0.96
+0.2	2	3.24	1.24	0.605	0.581	0.212	0.94	5.58	0.95
+0.1	1	2.56	1.56	0.599	0.587	0.106	0.94	6.07	0.94
0	0	2	2	0.593	0.593	0	—	6.25	0.94
–0.1	–1	1.56	2.56	0.587	0.599	–0.106	0.94	6.07	0.94
–0.2	–2	1.24	3.24	0.581	0.605	–0.212	0.94	5.58	0.95
–0.5	–5	0.70	5.70	0.567	0.619	–0.526	0.95	4.03	0.96
–1.0	–10	0.39	10.39	0.552	0.634	–1.041	0.96	2.32	0.98
–5.0	–50	0.08	50.08	0.513	0.673	–5.08	0.98	0.50	1.00
–10.0	–100	0.04	100.04	0.495	0.691	–10.1	0.99	0.25	1.00

The table also provides values for the dc gain v_O/v_I as well as the incremental gain v_o/v_i at the various values of v_O. The incremental gain is computed as follows

$$\frac{v_o}{v_i} = \frac{R_L}{R_L + R_{out}}$$

where R_{out} is the small-signal output resistance of the amplifier, given by Eq. (12.31). The incremental gain is the slope of the voltage-transfer characteristic, and the magnitude of its variation over the range of v_O is an indication of the linearity of the output stage. Observe that for $0 \leq |v_O| \leq 10$ V, the incremental gain changes from 0.94 to 1.00, about 6%. This is worse than the 1% we calculated for the class A stage in Example 12.1, but much better than a class B stage where the incremental gain becomes zero around $v_O =$ 0 V. Moreover, the class AB stage has much better power efficiency than the class A stage. Also notice that as v_O becomes positive, Q_N supplies more and more of i_L and the current in Q_P is correspondingly reduced. The opposite happens for negative v_O.

EXERCISE

12.5 To increase the linearity of the class AB output stage, the quiescent current I_Q is increased. The price paid is an increase in quiescent power dissipation. For the output stage considered in Example 12.4:
(a) Find the quiescent power dissipation.
(b) If I_Q is increased to 10 mA, find v_o/v_i at $v_O = 0$ and at $|v_O| = 10$ V, and hence the percentage change. Compare to the case in Example 12.4.
(c) Find the quiescent power dissipation for the case in (b).

Ans. (a) 60 mW; (b) 0.988 to 1.00; for a change of 1.2% compared to the 6% change in Example 12.4; (c) 300 mW

12.5 Biasing the Class AB Circuit

In this section we discuss approaches for generating the voltage V_{BB} and supplying the base currents required by the class AB output stage.

12.5.1 Biasing Using Diodes

Figure 12.13 shows a class AB circuit in which the bias voltage V_{BB} is generated by passing a constant current I_{BIAS} through a pair of diodes, D_1 and D_2. Typically, the output transistors are large devices capable of supplying large amounts of power. The biasing diodes, however, need not be large. Assuming the saturation current of the biasing diodes is $1/n$ times that of the output devices, the quiescent current in $Q_{N,P}$ will be $I_Q = nI_{\text{BIAS}}$. In an integrated circuit, we can set the ratio n precisely by realizing D_1 and D_2 as diode-connected transistors having the same construction as $Q_{N,P}$, but $1/n$ the emitter-junction area. This allows I_{BIAS} to be relatively small, saving power and improving the stage efficiency.

When the output stage of Fig. 12.13 is sourcing current to the load, the base current of Q_N increases from I_Q/β_N (which is usually small) to approximately i_L/β_N. This base current drive must be supplied by the current source I_{BIAS}. It follows that I_{BIAS} must be greater than the maximum anticipated base drive for Q_N. This sets a lower limit on the value of I_{BIAS}. Now, since $n = I_Q/I_{BIAS}$, and since I_Q is usually much smaller than the peak load current ($<10\%$), we see that we cannot make n a large number. In other words, we cannot make the diodes much smaller than the output devices. This is a disadvantage of the diode biasing scheme.

From the discussion above we see that the current through the biasing diodes will decrease when the output stage is sourcing current to the load. Thus the bias voltage V_{BB} will also decrease, and we must modify the analysis of Section 12.4 to take this into account.

The diode biasing arrangement has an important advantage: It can provide thermal stabilization of the quiescent current in the output stage. To appreciate this point, recall that the class AB output stage dissipates power under quiescent conditions. Power dissipation raises the internal temperature of the BJTs. From Chapter 6 we know that a rise in transistor temperature causes a decrease in its V_{BE} (approximately –2 mV/°C) if the collector current is held constant. Alternatively, if V_{BE} is held constant and the temperature increases, the

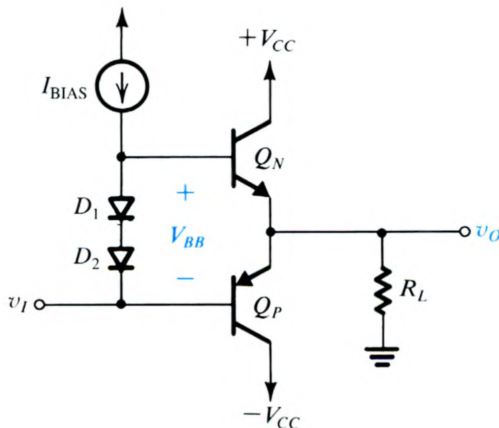


Figure 12.13 A class AB output stage using diodes for biasing. If the junction area of the output devices, Q_N and Q_P, is n times that of the biasing devices D_1 and D_2, a quiescent current $I_Q = nI_{\text{BIAS}}$ flows in the output devices.

collector current increases. The increase in collector current increases the power dissipation, which in turn increases the junction temperature and hence, once more, the collector current. So we have a positive-feedback mechanism that can cause **thermal runaway**. Unless checked, thermal runaway can lead to the ultimate destruction of the BJT. Diode biasing can be arranged to provide a compensating effect that can protect the output transistors against thermal runaway under quiescent conditions. Specifically, if the diodes are in close thermal contact with the output transistors, their temperature will increase by the same amount as Q_N and Q_P. Thus V_{BB} will decrease at the same rate as $V_{BEN} + V_{EBP}$, with the result that I_Q remains constant. We can easily achieve close thermal contact in IC fabrication. In discrete circuits, we can achieve it by mounting the bias diodes on the metal case of Q_N or Q_P. Finally, it is important to note that thermal runaway does not occur in MOS circuits.

Example 12.5

Consider the class AB output stage under the conditions that $V_{CC} = 15$ V, $R_L = 100\ \Omega$, and the output is sinusoidal with a maximum amplitude of 10 V. Let Q_N and Q_P be matched with $I_S = 10^{-13}$ A and $\beta = 50$. Assume that the biasing diodes have one-third the junction area of the output devices. Find the value of I_{BIAS} that guarantees a minimum of 1 mA through the diodes at all times. Determine the quiescent current and the quiescent power dissipation in the output transistors (i.e., at $v_O = 0$). Also find V_{BB} for $v_O = 0$, +10 V, and −10 V.

Solution

The maximum current through Q_N is approximately equal to $i_{L\max} = 10\text{ V}/0.1\text{ k}\Omega = 100$ mA. Thus the maximum base current in Q_N is approximately 2 mA. To maintain a minimum of 1 mA through the diodes, we select $I_{\text{BIAS}} = 3$ mA. The area ratio of 3 yields a quiescent current of 9 mA through Q_N and Q_P. The quiescent power dissipation is

$$P_{DQ} = 2 \times 15 \times 9 = 270\text{ mW}$$

For $v_O = 0$, the base current of Q_N is $9/51 \simeq 0.18$ mA, leaving a current of $3 - 0.18 = 2.82$ mA to flow through the diodes. Since the diodes have $I_S = \frac{1}{3} \times 10^{-13}$ A, the voltage V_{BB} will be

$$V_{BB} = 2V_T \ln\left(\frac{2.82\text{ mA}}{I_S}\right) = 1.26\text{ V}$$

At $v_O = +10$ V, the current through the diodes will decrease to 1 mA, resulting in $V_{BB} \simeq 1.21$ V. At the other extreme of $v_O = -10$ V, Q_N will conduct a very small current; thus its base current will be negligibly small and all of I_{BIAS} (3 mA) flows through the diodes, resulting in $V_{BB} \simeq 1.26$ V.

EXERCISES

12.6 For the circuit of Example 12.5, find i_N and i_P for $v_O = +10$ V and $v_O = -10$ V. (*Hint*: Use the V_{BB} values found in Example 12.5.)
Ans. 100.1 mA, 0.1 mA; 0.8 mA, 100.8 mA

12.7 If the collector current of a transistor is held constant, its v_{BE} decreases by 2 mV for every 1°C rise in temperature. Alternatively, if v_{BE} is held constant, then i_C increases by approximately $g_m \times 2$ mV for every 1°C rise in temperature. For a device operating at $I_C = 10$ mA, find the change in collector current resulting from an increase in temperature of 5°C.
Ans. 4 mA

12.5.2 Biasing Using the V_{BE} Multiplier

An alternative biasing arrangement that provides the designer with considerably more flexibility in both discrete and integrated designs is shown in Fig. 12.14. It consists of transistor Q_1 with a resistor R_1 connected between base and emitter and a feedback resistor R_2 connected between collector and base. We feed the resulting two-terminal network with a constant-current source I_{BIAS}. If we neglect the base current of Q_1, then R_1 and R_2 will carry the same current I_R, given by

$$I_R = \frac{V_{BE1}}{R_1} \tag{12.32}$$

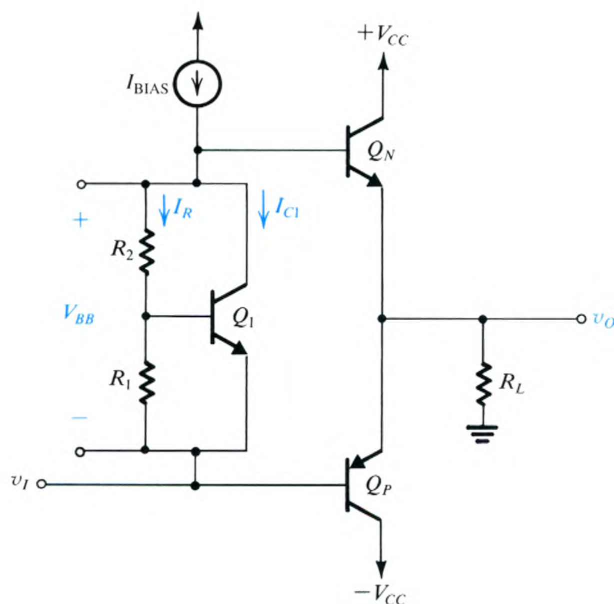


Figure 12.14 A class AB output stage using a V_{BE} multiplier for biasing.

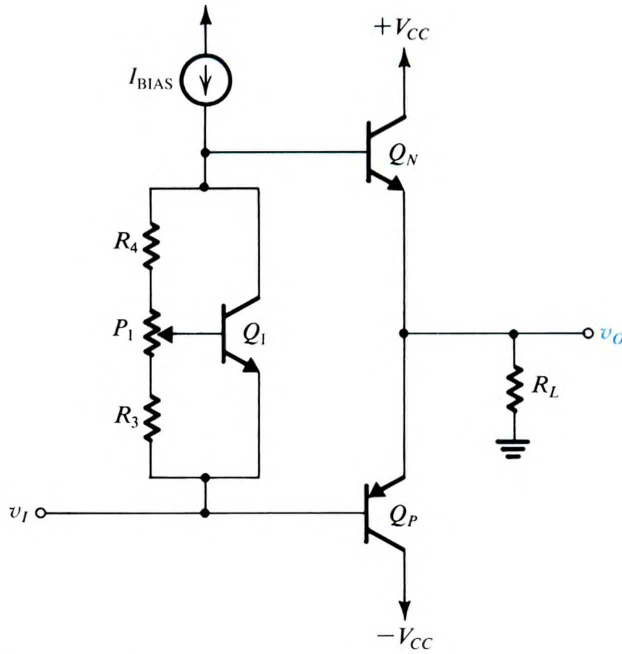


Figure 12.15 A discrete-circuit class AB output stage with a potentiometer used in the V_{BE} multiplier. We adjust the potentiometer to give the desired value of quiescent current in Q_N and Q_P.

and the voltage V_{BB} across the bias network will be

$$V_{BB} = I_R(R_1 + R_2)$$

$$= V_{BE1}\left(1 + \frac{R_2}{R_1}\right) \tag{12.33}$$

Thus the circuit simply multiplies V_{BE1} by the factor $(1 + R_2/R_1)$ and is known as the "V_{BE} multiplier." The multiplication factor is obviously under the designer's control and can be used to establish the value of V_{BB} needed to yield a desired quiescent current I_Q. In IC design it is relatively easy to accurately control the ratio of two resistances. In discrete-circuit design, we can use a potentiometer, as shown in Fig. 12.15, and manually set it to produce the desired value of I_Q.

The value of V_{BE1} in Eq. (12.33) is determined by the portion of I_{BIAS} that flows through the collector of Q_1; that is,

$$I_{C1} = I_{BIAS} - I_R \tag{12.34}$$

$$V_{BE1} = V_T \ln \frac{I_{C1}}{I_{S1}} \tag{12.35}$$

where we have neglected the base current of Q_N, which is normally small both under quiescent conditions and when the output voltage is swinging negative. However, for positive v_O, especially at and near its peak value, the base current of Q_N can become sizable and will

reduce the current available for the V_{BE} multiplier. Nevertheless, since large changes in I_{C1} correspond to only small changes in V_{BE1}, the decrease in current will be mostly absorbed by Q_1, leaving I_R, and hence V_{BB}, almost constant.

EXERCISE

12.8 Consider a V_{BE} multiplier with $R_1 = R_2 = 1.2\,\text{k}\Omega$, utilizing a transistor that has $V_{BE} = 0.6$ V at $I_C = 1$ mA, and a very high β. (a) Find the value of the current I that should be supplied to the multiplier to obtain a terminal voltage of 1.2 V. (b) Find the value of I that will result in the terminal voltage changing (from the 1.2-V value) by +50 mV, +100 mV, +200 mV, –50 mV, –100 mV, –200 mV.

Ans. (a) 1.5 mA; (b) 3.24 mA, 7.93 mA, 55.18 mA, 0.85 mA, 0.59 mA, 0.43 mA

Like the diode biasing network, the V_{BE}–multiplier circuit can provide thermal stabilization of I_Q. This is especially true if $R_1 = R_2$, and Q_1 is in close thermal contact with the output transistors.

Example 12.6

We need to redesign the output stage of Example 12.5 using a V_{BE} multiplier for biasing. Use a small-geometry transistor for Q_1 with $I_S = 10^{-14}$ A and design for a quiescent current $I_Q = 2$ mA.

Solution

Since the peak positive current is 100 mA, the base current of Q_N can be as high as 2 mA. We will therefore select $I_{\text{BIAS}} = 3$ mA, providing the multiplier with a minimum current of 1 mA.

Under quiescent conditions ($v_O = 0$ and $i_L = 0$) the base current of Q_N can be neglected and all of I_{BIAS} flows through the multiplier. We now must decide how this current (3 mA) is to be divided between I_{C1} and I_R. If we select I_R greater than 1 mA, the transistor will be almost cut off at the positive peak of v_O. Therefore, we will select $I_R = 0.5$ mA, leaving 2.5 mA for I_{C1}.

To obtain a quiescent current of 2 mA in the output transistors, V_{BB} should be

$$V_{BB} = 2V_T \ln \frac{2 \times 10^{-3}}{10^{-13}} = 1.19\text{ V}$$

We can now find $R_1 + R_2$ as follows:

$$R_1 + R_2 = \frac{V_{BB}}{I_R} = \frac{1.19}{0.5} = 2.38\text{ k}\Omega$$

At a collector current of 2.5 mA, Q_1 has

$$V_{BE1} = V_T \ln \frac{2.5 \times 10^{-3}}{10^{-14}} = 0.66\text{ V}$$

Example 12.6 *continued*

We can now find the value of R_1 as

$$R_1 = \frac{0.66}{0.5} = 1.32\ \text{k}\Omega$$

and R_2 as

$$R_2 = 2.38 - 1.32 = 1.06\ \text{k}\Omega$$

12.5.3 Use of Input Emitter Followers

Figure 12.16 shows a class AB circuit biased using transistors Q_1 and Q_2, which also function as emitter followers, providing the circuit with a high input resistance. In effect, the Q_1–Q_2 circuit functions as a unity-gain buffer amplifier. Since all four transistors are usually matched, and neglecting the effect of R_3 and R_4, we see that the quiescent current ($v_I = 0$, $R_L = \infty$) in Q_3 and Q_4 is equal to that in Q_1 and Q_2. Resistors R_3 and R_4 are usually

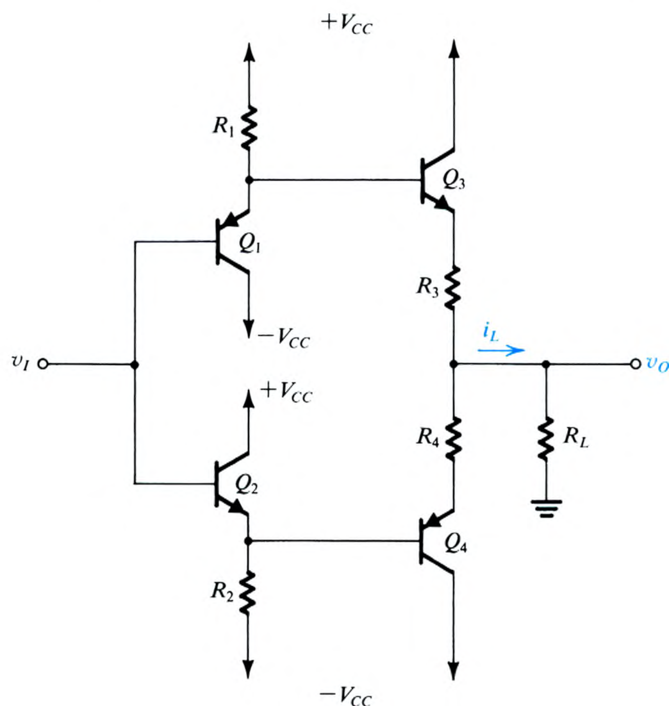


Figure 12.16 A class AB output stage with an input buffer. In addition to providing a high input resistance, the buffer transistors Q_1 and Q_2 bias the output transistors Q_3 and Q_4.

very small and are included to compensate for possible mismatches between Q_3 and Q_4. They also guard against the possibility of thermal runaway due to temperature differences between the input- and output-stage transistors. On this last point, note that an increase in the current of, say, Q_3 causes an increase in the voltage drop across R_3 and a corresponding decrease in V_{BE3}. Thus R_3 provides negative feedback that helps stabilize the current through Q_3.

EXERCISE

12.9 (*Note:* Although rather long, this exercise is very instructive.) Consider the circuit of Fig. 12.16 with $R_1 = R_2 = 5\ \text{k}\Omega$, $R_3 = R_4 = 0\ \Omega$, and $V_{CC} = 15$ V. Let the transistors be matched with $I_S = 3.3 \times 10^{-14}$ A and $\beta = 200$. (a) For $v_I = 0$ and $R_L = \infty$, find the quiescent current in each of the four transistors and v_O. (b) For $R_L = \infty$, find i_{C1}, i_{C2}, i_{C3}, i_{C4}, and v_O for $v_I = +10$ V and -10 V. (c) Repeat (b) for $R_L = 100\ \Omega$.

Ans. (a) 2.87 mA; 0 V; (b) for $v_I = +10$ V: 0.88 mA, 4.87 mA, 1.95 mA, 1.95 mA, +9.98 V; for $v_I = -10$ V: 4.87 mA, 0.88 mA, 1.95 mA, 1.95 mA, −9.98 V; (c) for $v_I = +10$ V: 0.38 mA, 4.87 mA, 100 mA, 0.02 mA, +9.86 V; for $v_I = -10$ V: 4.87 mA, 0.38 mA, 0.02 mA, 100 mA, −9.86 V

12.5.4 Use of Compound Devices

To increase the current gain of the output-stage transistors, and thus reduce the required base current drive, we often use the Darlington configuration shown in Fig. 12.17 to replace the *npn* transistor of the class AB stage. The Darlington configuration is equivalent to a single *npn* transistor having $\beta \simeq \beta_1\beta_2$, but almost twice the value of V_{BE}.

The Darlington configuration can be also used for *pnp* transistors, and this is indeed done in discrete-circuit design. In IC design, however, the lack of good-quality *pnp* transistors prompted the use of the alternative compound configuration shown in Fig. 12.18. This compound device is equivalent to a single *pnp* transistor having $\beta \simeq \beta_1\beta_2$. When fabricated with standard IC technology, Q_1 is usually a lateral *pnp* having a low β ($\beta = 5 - 10$) and poor high-frequency response ($f_T \simeq 5$ MHz); see Appendix A and Appendix K. The compound device, although it has a relatively high equivalent β, still suffers from a poor high-frequency response. It also suffers from another problem: The feedback loop formed by Q_1 and Q_2 is prone to high-frequency oscillations (with frequency near f_T of the *pnp* device, i.e., about 5 MHz). Methods exist for preventing such oscillations. We studied feedback-amplifier stability in Chapter 11.

To illustrate the application of the Darlington configuration and of the compound *pnp*, we show in Fig. 12.19 an output stage using both. Class AB biasing is achieved using a V_{BE} multiplier. Note that the Darlington *npn* adds one more V_{BE} drop, and thus the V_{BE} multiplier is required to provide a bias voltage of about 2 V. We investigate the design of this class AB stage in Problem 12.40.

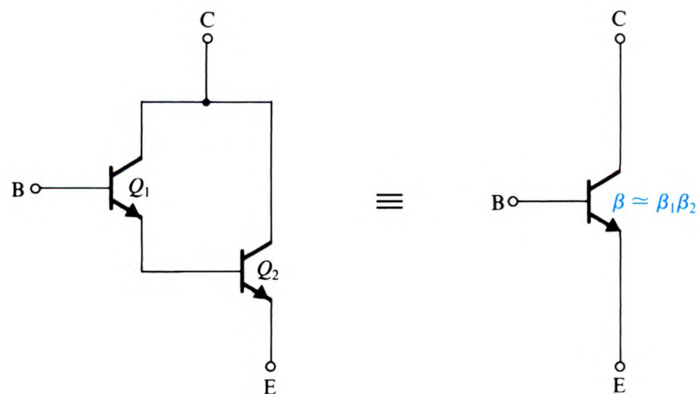


Figure 12.17 The Darlington configuration.

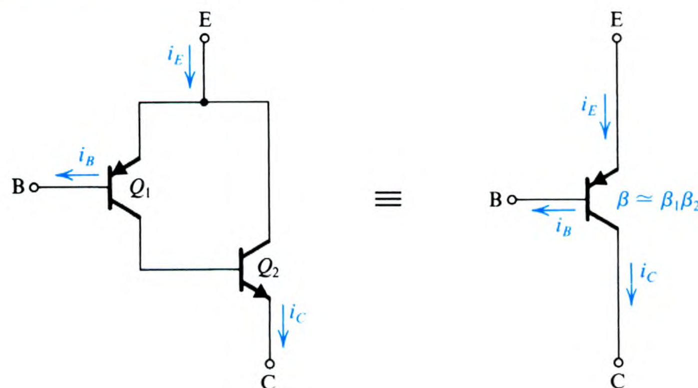


Figure 12.18 The compound-*pnp* configuration.

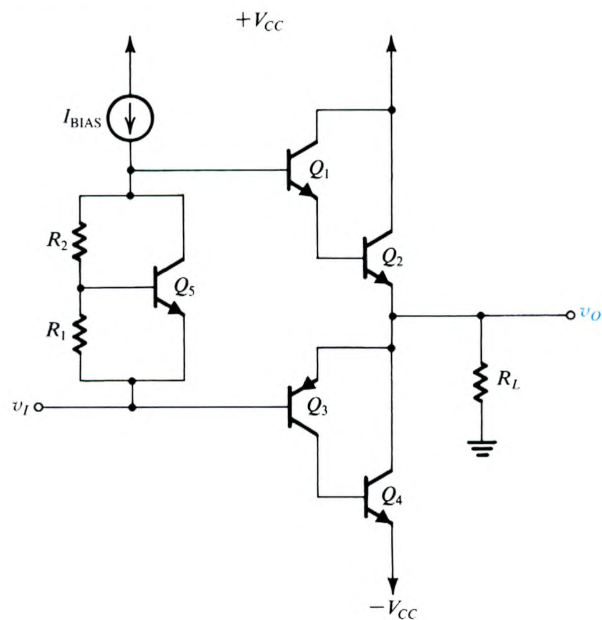


Figure 12.19 A class AB output stage using a Darlington *npn* and a compound *pnp*. We bias the circuit using a V_{BE} multiplier.

EXERCISE

12.10 (a) Refer to Fig. 12.18. Show that, for the composite *pnp* transistor,

$$i_B \simeq \frac{i_C}{\beta_N \beta_P}$$

and

$$i_E \simeq i_C$$

Hence show that

$$i_C \simeq \beta_N I_{SP} e^{v_{EB}/V_T}$$

and thus the transistor has an effective scale current

$$I_S = \beta_N I_{SP}$$

where I_{SP} is the scale current of the *pnp* transistor Q_1.
(b) For $\beta_P = 20$, $\beta_N = 50$, $I_{SP} = 10^{-14}$ A, find the effective current gain of the compound device and its v_{EB} when $i_C = 100$ mA.
Ans. (b) 1000; 0.651 V

12.6 CMOS Output Stages

We begin this section studying the source follower, essentially a CMOS variant of the BJT class A stage studied in Section 12.2. We will see that MOSFET device characteristics

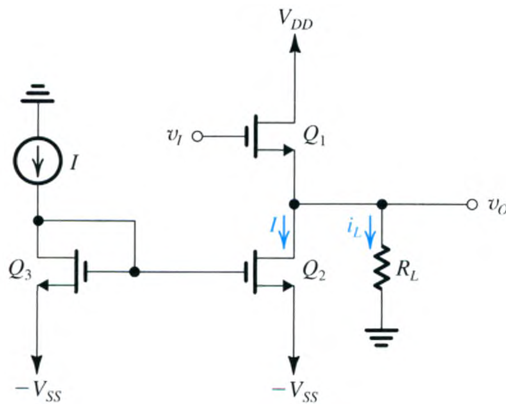


Figure 12.20 The source-follower (Q_1) output stage biased by an NMOS current mirror ($Q_{2,3}$).

impose significant limitations on the performance of these stages compared to their BJT counterparts. We will then look at attractive alternatives better suited to CMOS technologies.

12.6.1 The Source Follower

Figure 12.20 shows a CMOS variant of the class A stage where Q_1 is biased with a current I supplied by the current mirror $Q_{2,3}$. Transistor Q_1 will remain in saturation so long as

$$v_{O\max} = V_{DD} - v_{OV1} \tag{12.36}$$

As with the emitter follower, $v_{O\max}$ will be lower than Eq. (12.36) if v_I is restricted below V_{DD} by the preceding stage. To keep Q_2 in saturation,

$$v_{O\min} = -V_{SS} + V_{OV2} \tag{12.37}$$

Finally, as in the emitter follower, in order to maximize output swing the bias current I must be sufficient to source all of the load current,

$$I \geq \frac{-v_{O\min}}{R_L} = \frac{V_{SS} - V_{OV2}}{R_L} \tag{12.38}$$

Most likely, $v_{OV} > V_{CEsat}$, so that Eqs. (12.36) and (12.37) impose narrower limits on output swing than Eqs. (12.2) and (12.4) impose on the BJT emitter follower.

Furthermore, we may write,

$$v_O = v_I - v_{GS1} = v_I - V_{tn1} - v_{OV1} \tag{12.39}$$

Thus, the input waveform is faithfully reproduced at the output only insofar as the values of V_{tn1} and v_{OV1} are constant. Unfortunately, both may vary significantly over the full range of v_O. The threshold voltage V_{tn1} varies nonlinearly with $v_{S1} = v_O$ due to the body effect. This can be mitigated if the body of Q_1 is shorted to its source. Additionally, v_{OV1} varies with changes in $i_{D1} = I + i_L = I + v_O/R_L$, giving rise to distortion. In Example 12.7 we consider the distortion of the source follower by performing a small-signal analysis to find its incremental gain at several operating points, similar to what was done in Examples 12.1 and 12.4 for the BJT class A and class AB stages.

Example 12.7

In Fig. 12.20 let $V_{DD} = V_{SS} = 15$ V, $I = 120$ mA, and $R_L = 100\ \Omega$. Find the incremental (small-signal) gain of the circuit at $v_O = +10$ V, 0 V, and -10 V assuming $k_n = 300$ mA/V^2 and $V_{tn} = 0.8$ V. Neglect the output resistances $r_{o1,2}$ and assume the body of Q_1 is connected to its source.

Solution

A small-signal analysis of the source follower in Chapter 7 showed

$$A_v = \frac{R_L}{R_L + 1/g_{m1}} = \frac{1}{1 + 1/g_{m1}R_L}$$

The equations required to find the small-signal gain, along with the results at $v_O = +10$ V, 0 V, and -10 V, are shown in the table below:

v_O	$i_L = v_O/R_L$	$i_{D1} = I + i_L$	$1/g_{m1} = 1/\sqrt{2k_n I_{D1}}$	A_v
+10 V	100 mA	220 mA	2.75 Ω	0.973 V/V
0 V	0 mA	120 mA	3.72 Ω	0.964 V/V
−10 V	−100 mA	20 mA	9.13 Ω	0.916 V/V

This 6% variation in A_v is much larger than the 1% we observed in the emitter follower of Example 12.1, which consumes the same power. Hence, we can expect more distortion from the source follower.

The increased distortion of a source follower compared to an emitter follower may be explained by the generally lower small-signal transconductance of a MOSFET compared to a BJT conducting similar current. This, in turn, leads to a higher output resistance for the stage, which in Example 12.7 varies significantly compared to the load resistance. The reduced g_{m1} also makes the source follower gain more sensitive to variation in the device parameters that arise, for example, due to temperature changes.

Replacing the BJTs with MOSFETs in the class AB stage of Section 12.4 will cause similar limitations. Assuming the MOSFET v_{OV} is greater than BJT V_{CEsat}, the output swing will be reduced, and for the same power consumption, distortion will be higher due to the lower g_m of MOSFETs.

EXERCISES

12.11 The source follower in Fig. 12.20 operates from $V_{DD} = V_{SS} = 2.5$ V with $k_n = 50$ mA/V^2 and $V_{tn} = 0.5$ V. If $I = 10$ mA and $R_L = 200\ \Omega$, what input voltage v_I will give rise to $v_O = 0$ V? What is the minimum load resistance for which Eq. (12.38) remains satisfied?
Ans. 1.13 V; 187 Ω

12.12 In Fig. 12.20, $V_{DD} = V_{SS} = 3$ V with $k_n = 80$ mA/V^2, $V_{tn} = 0.5$ V, $I = 50$ mA, and $R_L = 100\ \Omega$. Find the incremental gain at $v_O = 0$ V and at v_{Omin}. What is the percentage change?
Ans. 0.899 V/V; 0.884 V/V; 1.7 %

12.6.2 An Alternative Using a Common-Source Transistor

Achieving v_{Omax} in Eq. (12.36) for Fig. 12.20 requires v_I to exceed V_{DD}. To alleviate this challenge, a common source configuration can be used instead, as shown in Fig. 12.21(a). Unfortunately, it has a gain of approximately $g_{m1}R_L$, which varies considerably with i_L, causing distortion.

The circuit is improved using a feedback amplifier with gain μ as shown in Fig. 12.21(b). A negative small-signal gain due to common-source Q_1 appears from the amplifier

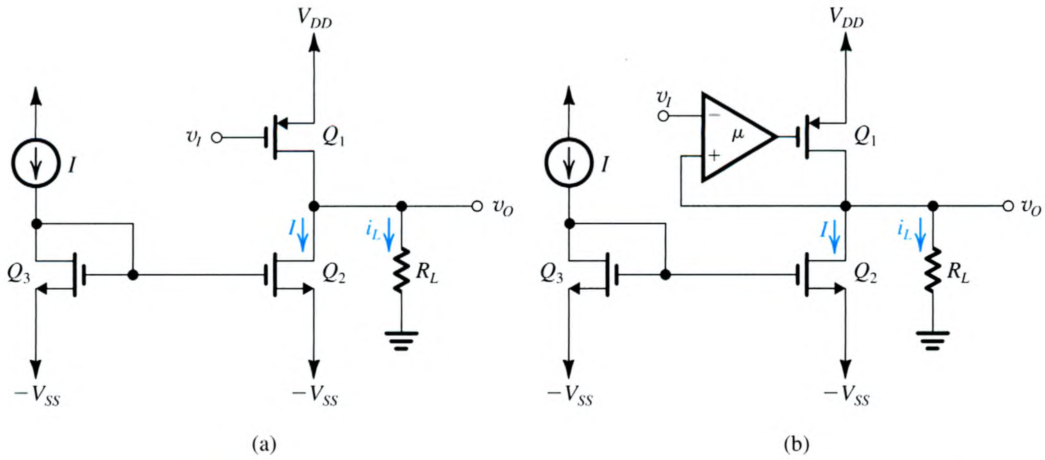


Figure 12.21 (**a**) A common-source circuit offers more swing at v_O. (**b**) Introducing an amplifier with negative feedback reduces the output resistance and keeps $v_O \cong v_I$.

output to v_O, so connecting v_O back to the amplifier's positive input terminal completes a negative feedback loop. The series-shunt feedback is appropriate for a voltage amplifier. It keeps the incremental closed-loop gain close to unity regardless of v_O, thus reducing distortion significantly. This is still considered a class A stage since current flows through Q_1 continuously, just as in the source and emitter followers.

Output Resistance Fig. 12.22(a) shows the equivalent circuit for finding the output resistance. We are interested in incremental quantities, so we have replaced V_{DD} with a short circuit to ground. Transistor Q_2 in Fig. 12.21 does not appear in Fig. 12.22(a) because we assume $r_{o2} \gg R_L$ and may therefore be ignored. Since the output is connected directly back to the input, the feedback factor is $\beta = 1$. Including the loading effects of the feedback network results in the A circuit shown in Fig. 12.22(b), with a gain given by

$$A \equiv \frac{v_o}{v_i} = \mu g_{mp}(r_{op}||R_L) \tag{12.40}$$

The values of the small-signal parameters g_{mp} and r_{op} are evaluated at the current at which Q_P is operating. The open-loop output resistance R_o is found by inspection as

$$R_o = R_L \parallel r_{op} \tag{12.41}$$

The output resistance with feedback R_{of} can now be found as

$$R_{of} = \frac{R_o}{1+A\beta} = \frac{(R_L \parallel r_{op})}{1+\mu g_{mp}(r_{op} \parallel R_L)} \tag{12.42}$$

and the output resistance R_{outp} is found by excluding R_L from R_{of}, that is,

$$R_{\text{outp}} = 1\Big/\left(\frac{1}{R_{of}} - \frac{1}{R_L}\right) \tag{12.43}$$

which results in

$$R_{\text{outp}} = r_{op} \Bigg\| \frac{1}{\mu g_{mp}} \simeq \frac{1}{\mu g_{mp}} \tag{12.44}$$

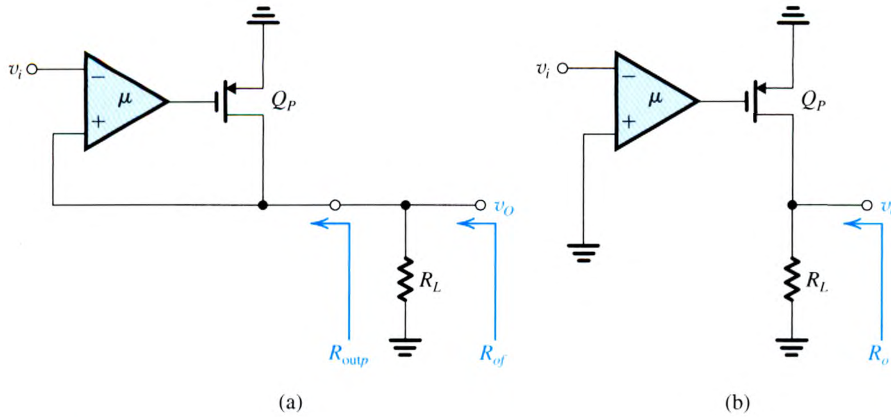


Figure 12.22 Determining the output resistance. (**a**) Defining the closed-loop resistances R_{of} and R_{outp}. (**b**) The circuit for finding the open-loop output resistance R_o.

Equation (12.44) can be lower than the output resistance of the source follower. In fact, we can make R_{outp} so low that even when it varies, the closed-loop gain remains very close to unity, thus ensuring low distortion.

We can also use the circuit in Fig. 12.21(b) as a voltage regulator, which is essentially a dc power amplifier intended to efficiently maintain a stable and precise constant voltage across a load. In such applications distortion is of no concern since v_I is constant. However, we still need low output resistance so that v_O remains constant in the presence of varying load current i_L.

A class AB variant of the circuit in Fig. 12.21(b) is shown in Fig. 12.23(a). The amplifiers are designed so that with $v_I = 0$ V, as shown in Fig. 12.23(b), their outputs are the gate voltages required to established a desired quiescent current I_Q. The gate voltages undergo incremental changes as v_I and v_O swing away from 0 V so that i_{DP} and i_{DN} source and sink the load current, respectively. To obtain class AB operation, we usually select I_Q to be 10% or so of the maximum output current. The output resistance of the bottom half of the circuit is similar to that of the top and appears in parallel. Thus, using Eq. (12.44)

$$R_{out} \simeq 1/\mu(g_{mp} + g_{mn}) \tag{12.45}$$

Usually the two transistors are matched,

$$k_p'\left(\frac{W}{L}\right)_p = k_n'\left(\frac{W}{L}\right)_n = k$$

Thus they have the same overdrive voltage, V_{OV}, at the quiescent point.

Consider the situation with v_I applied, illustrated in Fig. 12.23(a). The voltage at the output of each of the error amplifiers increases by $\mu(v_O - v_I)$. Thus v_{SGP} decreases by $\mu(v_O - v_I)$ and v_{GSN} increases by $\mu(v_O - v_I)$, and we can write

$$i_{DP} = \frac{1}{2}k[V_{OV} - \mu(v_O - v_I)]^2 = \frac{1}{2}kV_{OV}^2\left[1 - \mu\frac{v_O - v_I}{V_{OV}}\right]^2$$

$$= I_Q\left(1 - \mu\frac{v_O - v_I}{V_{OV}}\right)^2 \tag{12.46}$$

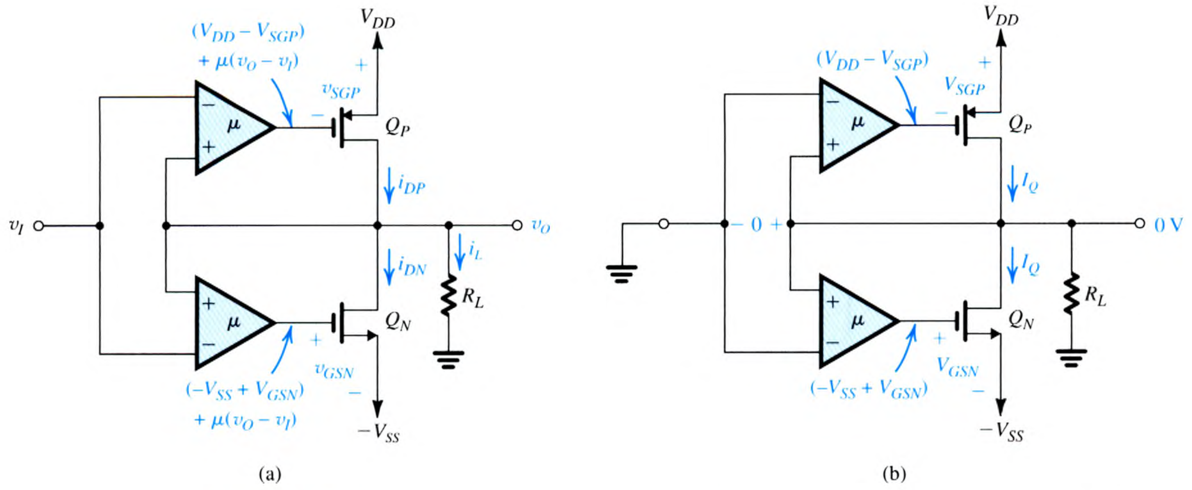


Figure 12.23 **(a)** A CMOS class AB common-source output stage using feedback. **(b)** Quiescent conditions.

Assuming the two transistors are matched to have the same quiescent overdrive voltage, V_{OV},

$$i_{DN} = I_Q\left(1 + \mu\frac{v_O - v_I}{V_{OV}}\right)^2 \tag{12.47}$$

At the output node we have

$$i_L = i_{DP} - i_{DN} \tag{12.48}$$

Substituting for $i_L = v_O/R_L$ and for i_{DP} and i_{DN} from Eqs. (12.46) and (12.47), and solving the resulting equation to obtain v_O, results in

$$v_O = \frac{v_I}{1 + \dfrac{V_{OV}}{4\mu I_Q R_L}} \tag{12.49}$$

Since $g_{mn} = g_{mp} = g_m = 2I_Q/V_{OV}$, using Eq. (12.45)

$$R_{out} = \frac{V_{OV}}{4\mu I_Q} \tag{12.50}$$

Combining Eqs. (12.49) and (12.50),

$$\frac{v_O}{v_I} = \frac{1}{1 + R_{out}/R_L} \tag{12.51}$$

Thus, selecting large values for μ and I_Q can reduce R_{out}, which keeps the gain close to unity and reduces crossover distortion. However, a large μ can make the stage sensitive to offset voltages that are inevitably present in the error amplifiers, and large I_Q implies high quiescent power dissipation.

EXERCISES

12.13 In the circuit in Fig. 12.21(b), $g_{mp} = 50$ mA/V. Find the gain μ required to ensure $R_{out} \simeq 1\ \Omega$. If $V_{OV1} = 300$ mV, what bias current I is required?
Ans. 20 V/V; 7.5 mA

12.14 The circuit of Fig. 12.23 delivers an output swing up to $|v_O| = 2$ V to a load $R_L = 100\ \Omega$. Find the quiescent current I_Q equal to 10% of the maximum output current. With $\mu = 10$ V/V and $V_{OV1,2} = 300$ mV, what is the corresponding output resistance at the quiescent point?
Ans. 2 mA; 3.75 Ω

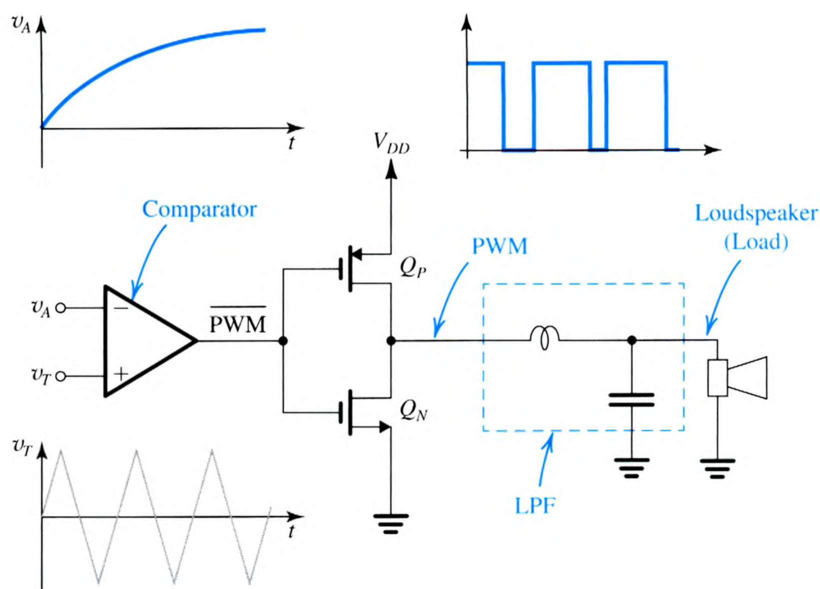
12.6.3 Class D Power Amplifiers

We previously mentioned that class D amplifiers operate by generating rectangular waveforms whose average value varies in proportion to the input signal. The main benefit of the class D stage is that its efficiency can approach 100% since the output transistors operate as switches. When turned on, an ideal switch has zero voltage drop, thus dissipating no power. When turned off, an ideal switch conducts zero current and still dissipates no power. By alternating between these two states, the output transistors in a class D stage can dissipate very little power and provide high efficiency.

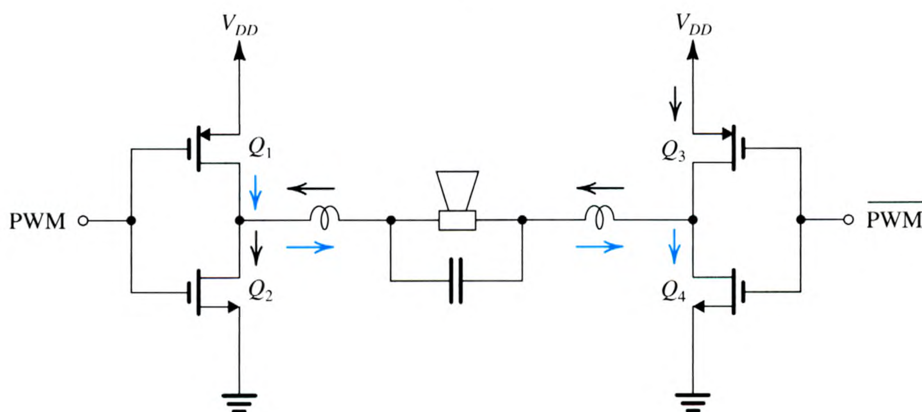
Generation of a pulsewidth modulated (PWM) waveform, suitable for a class D stage, is illustrated in Fig. 12.24(a), where the input, v_A, is compared with a triangular waveform, v_T. The comparison is performed by a *comparator* circuit, which produces a high output whenever v_T exceeds v_A, and a low output otherwise. This waveform is inverted by transistors $Q_{P,N}$ to produce the PWM waveform. The PWM frequency is the same as that of the triangular wave, f_s. The fraction of each cycle where the waveform is high is called its *duty ratio*, which is proportional to the corresponding instantaneous value of v_A. We then reconstruct the original waveform with a low-pass filter formed by the inductor and capacitor, which essentially provides a time-average of the PWM waveform.[3] To simplify the filter design, we choose f_s at least 10 times that of the highest frequency of v_A. The design of low-pass filters is studied in Chapter 14.

In Fig. 12.24(a), the two complementary MOS switches Q_P and Q_N connect the output node alternately to V_{DD} and ground, supplying the large current required by the low-resistance load. To double the voltage excursion across the load, we can use the scheme in Fig. 12.24(b). Here we're using both PWM and its logical inverse $\overline{\text{PWM}}$ in a differential driving arrangement. When PWM is high and thus $\overline{\text{PWM}}$ is low, Q_3 and Q_2 are turned on while Q_1

[3]The PWM signal has frequency components at f_s and its harmonics, all of which will be higher than the frequency of the input signal and thus can be easily removed by the low-pass filter.



(a)



(b)

Figure 12.24 Two schemes for driving the load of a class D amplifier. The differential scheme in (**b**) results in doubling the voltage excursion across the load.

and Q_4 are off. Thus current flows from V_{DD} to ground through the load (from right to left). The opposite happens when PWM is low. Thus the voltage across the load will be twice that obtained with the arrangement in Fig. 12.24(a). The circuit of the differential driving arrangement is known as an H bridge[4] and can result in a maximum sinusoidal output voltage of amplitude V_{DD}.

[4]The name H arises from the resemblance of the circuit diagram to the letter H: The switches represent the vertical strokes of H and the load with the filter represents the horizontal stroke.

HANS CAMENZIND: THE INVENTOR OF THE CLASS D AMPLIFIER

In 1966, while working for Mallory (now Duracell), Swiss-born engineer Hans Camenzind (1934–2012) filed a patent on "Pulse-Width Modulation Circuits," introducing the two-state or class D amplifier. Camenzind had begun his U.S. career in 1960, at Transitron, an early semiconductor manufacturer. Later, in 1969, while at Mallory, he filed a second patent on a "Two-State Amplifier." The class D amplifier has revolutionized the field of high-efficiency compact amplifiers. Today these devices, ubiquitous in mobile phones, also see application at enormous power levels for live-performance sound systems with thousands of watts per channel.

As mentioned previously, the power conversion efficiency of a class D stage is limited to about 85–90% by several practical limitations. These include the finite on-resistance of the switches, Q_{1-4}, power dissipation in the driving circuits required to turn the power switches on, and the momentary *shoot-through current* that arises during switching when current can pass directly from V_{DD} to ground. Nevertheless, class D amplifiers achieve efficiencies far superior to those of the class AB stage.

As a final note, distortion in the class D amplifier can be contributed by the PWM modulation scheme, by inaccuracy of the duty ratio at the output node due to finite switching speeds of the power MOSFETs, and by the quality factor of the L and C elements in the output filter. Distortion in power amplifiers is quantified by the **total harmonic distortion** (THD). This is the rms value of the harmonic components of the output signal, excluding the fundamental, expressed as a percentage of the rms value of the fundamental. Class D amplifiers typically exhibit THD of 0.1% to 1% at best. Another imperfection of class D amplifiers is the generation of unwanted switching noise, usually in the inaudible range as electromagnetic interference. As a result, class D amplifiers are most useful in applications where power-conversion efficiency is of paramount importance. Class A and AB amplifiers can achieve THD figures of less than 0.01% and are mostly used in high-fidelity applications.

EXERCISES

12.15 Consider the comparator in Fig. 12.24(a) with the triangular wave v_T having ± 10 V peak voltages and comparator output levels of ± 10 V. Find the duty ratio D and the average of the output voltage for the case in which v_A is a constant voltage of magnitude (a) 0 V; (b) +5 V; (c) +10 V; (d) -5 V; (e) -10 V

Ans. (a) 50%, 0 V; (b) 75%, +5 V; (c) 100%, +10 V; (d) 25%, -5 V; (e) 0 %; -10 V

12.16 If the audio signal v_A has a frequency spectrum of 20 Hz to 20 kHz, what is an appropriate value for f_s? Now if the low-pass filter is of second order, with its passband edge at 20 kHz and its gain falling off at 40 dB/decade, what is the attenuation encountered by the PWM component with frequency f_s?

Ans. 200 kHz; 40 dB

12.17 If the differential switching scheme shown in Fig. 12.24(b) is utilized and v_A is a sine wave, what is the maximum peak amplitude achieved across R_L and what is the maximum power delivered to R_L? Evaluate these quantities for $V_{DD} = 35$ V and $R_L = 8\ \Omega$. Now, if the power-conversion efficiency is 90%, what is the power delivered by the power supplies?

Ans. V_{DD}, $\frac{V_{DD}^2}{2R_L}$; 35 V, 76.6 W; 85.1 W

12.7 Power Transistors

BJTs and MOSFETs used to deliver load currents in the ampere range, support voltages around or above 100 V, and dissipate tens-of-watts of power are called **power transistors**. In this section we introduce the characteristics and specifications of power transistors.[5] Significantly more detail is available as ancillary material on the companion website.

12.7.1 Packages and Heat Sinks

Power transistors are basically larger versions of their small-signal counterparts; hence they retain similar characteristics. However, as will be explained shortly, their structures are modified for optimal voltage and current capabilities. Also, discrete power transistors are housed in special packages such as those in Fig. 12.25. The packages are usually mounted on **heat sinks**, special metal surfaces whose function is to facilitate the conduction of heat away from the transistor, thus keeping its internal temperature within safe operating limits. A typical heat sink is shown in Fig. 12.26.

12.7.2 Power BJTs

The parameters of power BJTs can differ somewhat from those of small-signal devices. Important differences include:

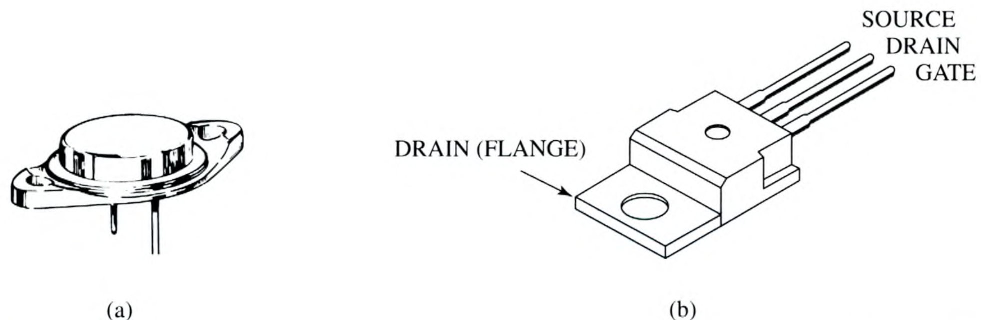


Figure 12.25 Most popular packages for power transistors: (**a**) TO-03 metal package; (**b**) TO-220 plastic package.

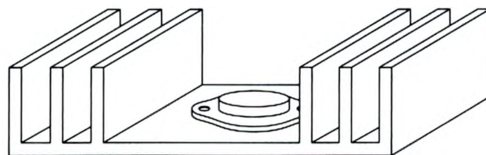


Figure 12.26 Typical heat sink.

[5] Other semiconductor devices used in power electronic applications include thyristors, silicon-controlled rectifiers (SCRs), and insulated-gate bipolar transistors (IGBTs). These are usually used in applications requiring much higher current and voltage ratings than those studied here.

1. The current gain β is low, typically in the range of 10 to 80, but can be as low as 5. Here it is important to recall that β is a function of current and has a positive temperature coefficient (refer to Fig. 6.34).
2. The maximum collector current $I_{C\max}$ is typically in the ampere range but can be as high as 100 A.
3. The breakdown voltage (BV_{CEO}; refer to Fig. 6.33) is typically 50 V to 100 V but can be as high as 500 V.
4. I_{CBO} is large (a few tens of microamps) and, as usual, doubles for every 10°C rise in temperature.
5. At high currents, r_π becomes small (a few ohms) and the extrinsic bulk resistance of the base region, denoted r_x, becomes important.
6. The transition frequency f_T is low (a few megahertz), C_μ is large (hundreds of picofarads), and C_π is even larger.
7. At high currents, the exponential i_C–v_{BE} relationship exhibits a factor-of-2 reduction in the exponent: that is, $i_C = I_S e^{v_{BE}/2V_T}$.

At large collector currents, the low β means we need a large base current. This can complicate the design of the circuit that drives the output transistors. We can use the Darlington configuration discussed in Section 12.5 to provide a higher effective β.

Power BJTs can suffer permanent damage if we operate them beyond any one of several physical limits:

1. Exceeding the maximum allowable current, $I_{C\max}$, can melt the wires that bond the device to the package terminals.
2. We must restrict the combination of collector current, i_C, and collector-emitter voltage, v_{CE}, within safe limits to keep the BJT from overheating. We can improve these limits by cooling the BJT case. For example, this can be done by blowing cool air over heat sinks.
3. Localized temperature increases can occur within the BJT structure itself because current flow across the emitter-base junction is not uniform, resulting in a **second-breakdown** limit.
4. Exceeding the collector-to-emitter breakdown voltage BV_{CEO} causes avalanche breakdown of the collector-base junction (Section 6.4).

These overlapping considerations allow us to define the BJT's safe operating area (SOA) as illustrated in Fig. 12.27. Such a plot often appears on power BJT data sheets.

12.7.3 Power MOSFETs

Power MOSFETs have gained popularity in the design of power amplifiers and output stages as a result of the following properties.

1. Unlike BJTs, MOSFETs do not require dc gate drive current. This greatly simplifies the design of the driving circuitry.
2. MOSFETs can operate at much higher switching speeds than BJTs, a definite advantage for power circuits employing switching, such as class D amplifiers.
3. The thermal characteristics of MOSFETs are superior to those of BJTs. They benefit from an extension of their SOA.

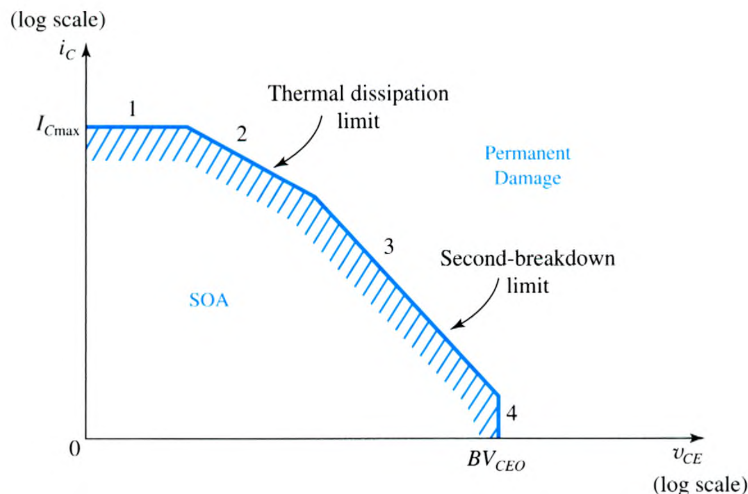


Figure 12.27 Safe operating area (SOA) of a BJT.

A particularly popular power MOSFET structure is the **double-diffused** or **DMOS** transistor. It has a radically different structure that simultaneously offers high current capability (50 A is possible) and very high breakdown voltage (as high as 600 V). Although it exhibits characteristics that are quite similar to those of the small-signal MOSFET studied in Chapter 5, some important differences exist.

Power MOSFETs have threshold voltages in the range of 1 V to 4 V. In saturation, the drain current is related to v_{GS} by the conventional square-law characteristic. However, the i_D–v_{GS} characteristic becomes linear for larger values of v_{GS}. The linear portion of the characteristic occurs as a result of the high electric field along the short channel, causing the velocity of charge carriers to reach an upper limit, a phenomenon known as **velocity saturation**.[6] The linear i_D–v_{GS} relationship implies a constant g_m in the velocity-saturation region.

Of considerable interest in the design of MOS power circuits is the variation of the MOSFET characteristics with temperature, illustrated in Fig. 12.28. Notice that there is a value of v_{GS} (in the range of 4 V to 6 V for most power MOSFETs) at which the temperature coefficient of i_D is zero. At higher values of v_{GS}, i_D exhibits a negative temperature coefficient. This is a significant property: It implies that a MOSFET operating beyond the zero-temperature-coefficient point does not risk the possibility of thermal runaway. This is *not* the case, however, at low currents (i.e., lower than the zero-temperature-coefficient point). In the (relatively) low-current region, the temperature coefficient of i_D is positive, and the power MOSFET can easily suffer thermal runaway (with unhappy consequences). Since class AB output stages are biased at low currents, we must do something to guard against thermal runaway.

The reason for the positive temperature coefficient of i_D at low currents is that $v_{OV} = (v_{GS} - V_t)$ is relatively low, and the temperature dependence is dominated by the negative temperature coefficient of V_t (in the range of –3 mV/°C to –6 mV/°C), which causes v_{OV} to rise with temperature.

[6]Velocity saturation occurs also in standard MOSFET structures when the channel length is in the submicron range. We will discuss velocity saturation in some detail in Section 17.4.

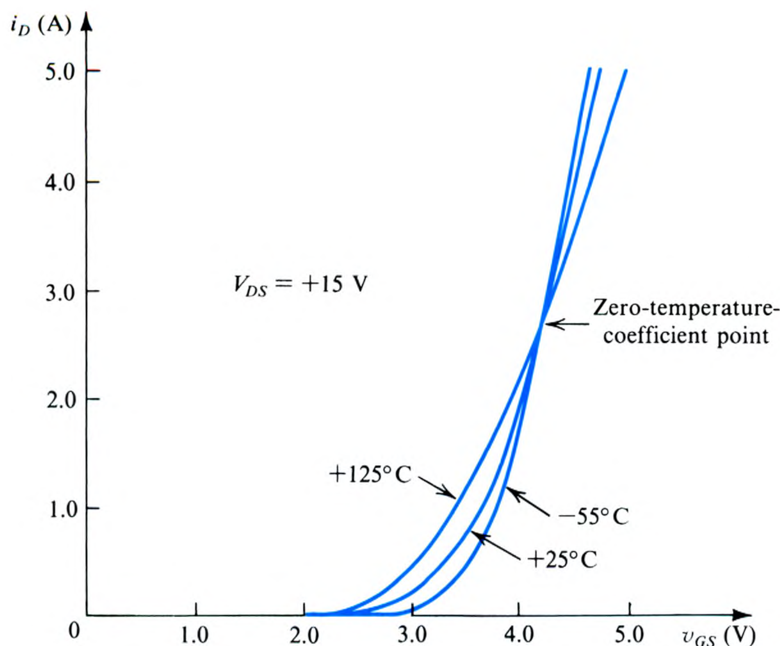


Figure 12.28 The i_D–v_{GS} characteristic curve of a power MOS transistor (IRF 630, Siliconix) at case temperatures of −55°C, +25°C, and +125°C. (Courtesy of Siliconix Inc.)

Summary

- Output stages are classified according to the transistor conduction angle: class A (360°), class AB (slightly more than 180°), class B (180°), and class C (less than 180°).
- The most common class A output stage is the emitter follower. It is biased at a current greater than the peak load current.
- The class A output stage dissipates its maximum power under quiescent conditions ($v_O = 0$). It achieves a maximum power-conversion efficiency of 25%.
- The class B stage is biased at zero current, and thus dissipates no power in quiescence.
- The class B stage can achieve a power-conversion efficiency as high as 78.5%. It dissipates its maximum power for $\hat{V}_o = (2/\pi)V_{CC}$.
- The class B stage suffers from crossover distortion.
- The class AB output stage is biased at a small current; thus both transistors conduct for small input signals, and there's virtually no crossover distortion.
- Except for an additional small quiescent power dissipation, the power relationships of the class AB stage are similar to those in class B.
- To guard against the possibility of thermal runaway, we make the bias voltage of the class AB circuit vary with temperature in the same manner as does V_{BE} of the output transistors.
- Use of the Darlington configuration in the class AB output stage reduces the base-current drive requirement. In integrated circuits, we commonly use the compound *pnp* configuration.
- A CMOS class A stage suffers from a higher output resistance than its BJT counterpart. Also, the output resistance can vary significantly over the full range of output voltages, resulting in distortion.
- The classical CMOS class AB output stage suffers from reduced output signal swing. We can overcome this problem by replacing the source-follower output transistors with a pair of complementary devices operating in the common-source configuration.
- The CMOS class AB output stage with common-source transistors allows the output voltage to swing to within an overdrive voltage from each of the two power supplies. Using amplifiers in the feedback path of each of the output transistors reduces both the output resistance and the distortion of the stage.

- Class D amplifiers convert the input signal into a pulsewidth-modulated (PWM) signal. The latter is then used to drive complementary MOS switches that supply the load with power. A low-pass filter is used to eliminate the high-frequency components introduced by the switching waveform. Power-conversion efficiences in the range of 85% to 90% are achieved.
- MOSFETs have gained popularity over BJTs in the design of high-power output stages. This is due to their higher speed of operation and to the fact that they do not need a steady supply of gate currents, which allows them to be used with relatively simple driving circuitry.
- The DMOS transistor is a MOSFET power device capable of both high-current and high-voltage operation.
- The drain current of a power MOSFET exhibits a positive temperature coefficient at low currents, and thus the device can suffer thermal runaway. At high currents the temperature coefficient of i_D is negative.

PROBLEMS

Computer Simulation Problems

SIM Problems identified by this icon are intended to demonstrate the value of using SPICE simulation to verify hand analysis and design, and to investigate important issues such as allowable signal swing and amplifier nonlinear distortion. Instructions to assist in setting up simulations for all the indicated problems can be found in the corresponding files on the website. Note that if a particular parameter value is not specified in the problem statement, you are to make a reasonable assumption.

Section 12.2: Class A Output Stage

12.1 A class A emitter follower, biased using the circuit shown in Fig. 12.3, uses $V_{CC}=10$ V, $R = R_L = 1$ kΩ, with all transistors (including Q_3) identical. Assume $V_{BE}=0.7$ V, $V_{CEsat}=0.3$ V, and β to be very large. For linear operation, what are the upper and lower limits of output voltage, and the corresponding inputs? How do these values change if the emitter–base junction area of Q_3 is made twice as big as that of Q_2? Half as big?

12.2 A source-follower circuit using NMOS transistors is constructed following the pattern shown in Fig. 12.3. All three transistors used are identical, with $V_t=0.5$ V and $\mu_n C_{ox} W/L = 20\,\text{mA/V}^2$; $V_{CC}=2.5$ V, $R = R_L = 1$ kΩ. For linear operation, what are the upper and lower limits of the output voltage, and the corresponding inputs?

D 12.3 Using the follower configuration shown in Fig. 12.3 with ±5-V supplies, provide a design capable of ±3-V outputs with a 200-Ω load, using the smallest possible total supply current. You are provided with four identical, high-β BJTs and a resistor of your choice. Select a standard resistor value of 5% tolerance, and specify the maximum power drawn from the negative supply.

D 12.4 An emitter follower using the circuit of Fig. 12.3, for which the output voltage range is ±5 V, is required using $V_{CC}=10$ V. The circuit is to be designed such that the current variation in the emitter-follower transistor is no greater than a factor of 10, for load resistances as low as 100 Ω. What is the value of R required? Find the incremental voltage gain of the resulting follower at $v_O=+5$, 0, and –5 V, with a 100-Ω load. What is the percentage change in gain over this range of v_O?

D 12.5 The class A emitter follower in Fig. 12.3 is to operate from ±10 V supplies and drive a 16-Ω load with a sinusoid with 5-V peak amplitude. If the incremental gain is to remain within 2% of unity, what must be the bias current I? What is the resulting efficiency?

***12.6** Consider the operation of the follower circuit of Fig. 12.3 for which $R_L = V_{CC}/I$, when driven by a square wave such that the output ranges from $+V_{CC}$ to $-V_{CC}$ (ignoring V_{CEsat}). For this situation, sketch the equivalent of Fig. 12.5 for v_O, i_{C1}, and p_{D1}. Repeat for a square-wave output that has peak levels of $\pm V_{CC}/2$. What is the average power dissipation

problems with blue numbers are considered essential; * = difficult problem; ** = more difficult; *** = very challenging
SIM = simulation; D = design problem; ● = see related video example

in Q_1 in each case? Compare these results to those for sine waves of peak amplitude V_{CC} and $V_{CC}/2$, respectively.

12.7 Consider the situation described in Problem 12.6. For square-wave outputs having $\pm V_{CC}$ levels and $\pm\frac{1}{2}V_{CC}$ levels, and for sine waves of the same peak-to-peak values, find the average power loss in the current-source transistor Q_2.

12.8 Reconsider the situation described in Exercise 12.3 for variation in V_{CC}—specifically for $V_{CC}=16$ V, 12 V, 10 V, and 8 V. Assume V_{CEsat} is nearly zero. What is the power-conversion efficiency in each case?

D 12.9 The emitter-follower output stage of Fig. 12.3 is designed to provide a maximum output swing of $\pm\hat{V}$ volts, across the load R_L. Neglecting the saturation voltage, what are the minimum required values of V_{CC} and I? Now, if the output voltage is a sine wave of peak amplitude $(\hat{V}/2)$, what is the power-conversion efficiency realized?

Section 12.3: Class B Output Stage

12.10 Consider the circuit of a complementary-BJT class B output stage. For what amplitude of input signal does the crossover distortion represent a 20% loss in peak amplitude?

12.11 Consider the feedback configuration with a class B output stage shown in Fig. P12.11. Let the amplifier gain $A_0 = 20$ V/V. Derive an expression for v_O versus v_I, assuming that $|V_{BE}| = 0.7$ V. Sketch the transfer characteristic v_O versus v_I, and compare it with that without feedback.

SIM **12.12** Consider the class B output stage, using MOSFETs, shown in Fig. P12.12. Let the devices have $|V_t| = 0.5$ V and $\mu C_{ox}W/L = 2\text{ mA/V}^2$. With a 10-kHz sine-wave input of 5-V peak and a high value of load resistance, what peak output would you expect? What fraction of the sine-wave period does the crossover interval represent? For what value of load resistor is the peak output voltage reduced to half the input?

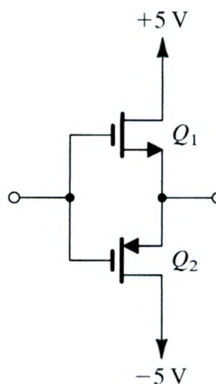


Figure P12.12

12.13 Consider the complementary-BJT class B output stage and neglect the effects of finite V_{BE} and V_{CEsat}. For ±5-V power supplies and an 8-Ω load resistance, what is the maximum sine-wave output power available? What supply power corresponds? What is the power-conversion efficiency? For output signals of half this amplitude, find the output power, the supply power, and the power-conversion efficiency.

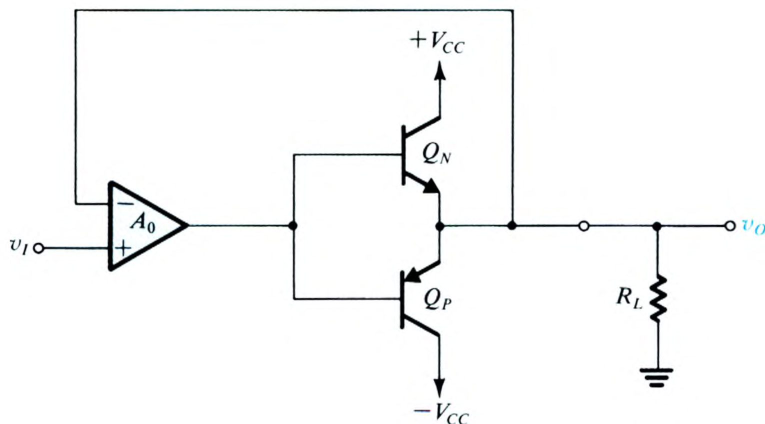


Figure P12.11

D 12.14 A class B output stage operates from ±10-V supplies. Assuming relatively ideal transistors, what is the output voltage for maximum power-conversion efficiency? What is the output voltage for maximum device dissipation? If each of the output devices is individually rated for 5-W dissipation, and a factor-of-2 safety margin is to be used, what is the smallest value of load resistance that can be tolerated, if operation is always at full output voltage? If operation is allowed at half the full output voltage, what is the smallest load permitted? What is the greatest possible output power available in each case?

D 12.15 A class B output stage is required to deliver an average power of 10 W into an 8-Ω load. The power supply should be 2 V greater than the corresponding peak sine-wave output voltage. Determine the power-supply voltage required (to the nearest volt in the appropriate direction), the peak current from each supply, the total supply power, and the power-conversion efficiency. Also, determine the maximum possible power dissipation in each transistor for a sine-wave input.

12.16 Consider the class B BJT output stage with a square-wave output voltage of amplitude $\hat{V}_o$ across a load R_L and employing power supplies $\pm V_{SS}$. Neglecting the effects of finite V_{BE} and V_{CEsat}, determine the load power, the supply power, the power-conversion efficiency, the maximum attainable power-conversion efficiency and the corresponding value of $\hat{V}_o$, and the maximum available load power. Also find the value of $\hat{V}_o$ at which the power dissipation in the transistors reaches its peak, and the corresponding value of power-conversion efficiency.

12.17 Sketch a graph for the small-signal voltage gain of the class B circuit of Fig. 12.6 as a function of v_I, for v_I both positive and negative.

Section 12.4: Class AB Output Stage

12.18 A class AB output stage, such as that in Fig. 12.10, utilizing transistors with $I_S = 10^{-15}$ A, is biased at a quiescent current $I_Q = 1$ mA. Find V_{BB}, the output resistance R_{out} at $v_I = 0$, and the corresponding small-signal voltage gain. The load resistance $R_L = 100\ \Omega$. What does the incremental gain become when $v_O = 10$ V?

D 12.19 Design the quiescent current of a class AB BJT output stage so that the incremental voltage gain for v_I in the vicinity of the origin is in excess of 0.97 V/V for loads larger than 50 Ω. Assume that the BJTs have V_{BE} of 0.7 V at a current of 100 mA and determine the value of V_{BB} required.

D 12.20 A class AB output stage, such as that in Fig. 12.10, drives a load resistance R_L of 100 Ω. What bias current I_Q will serve to limit the variation in the small-signal voltage gain to 2% as i_L changes from 0 to 50 mA?

12.21 For the class AB output stage considered in Example 12.4, add two columns to the table of results as follows: the total input current drawn from v_I (i_I, mA); and the large-signal input resistance $R_{in} \equiv v_I/i_I$. Assume $\beta_N = \beta_P = \beta = 49$. Compare the values of R_{in} to the approximate value obtained using the resistance-reflection rule, $R_{in} \simeq \beta R_L$.

12.22 In this problem we investigate an important trade-off in the design of the class AB output stage of Fig. 12.10: Increasing the quiescent current I_Q reduces the nonlinearity of the transfer characteristic at the expense of increased quiescent power dissipation. As a measure of nonlinearity, we use the maximum deviation from unity of the stage incremental gain, which occurs at $v_O = 0$, namely,

$$\epsilon = 1 - v_o/v_i\big|_{v_O=0}$$

(a) Show that ϵ is given by

$$\epsilon = \frac{V_T/2I_Q}{R_L + (V_T/2I_Q)}$$

which for $2I_QR_L \gg V_T$ can be approximated by

$$\epsilon \simeq V_T/2I_QR_L$$

(b) If the stage is operated from power supplies of $\pm V_{CC}$, find the quiescent power dissipation, P_D.

(c) Show that for given V_{CC} and R_L, the product of the quiescent power dissipation and the gain error is a constant given by

$$\epsilon P_D \simeq V_T\left(\frac{V_{CC}}{R_L}\right)$$

(d) For $V_{CC} = 10$ V and $R_L = 100\ \Omega$, find the required values of P_D and I_Q if ϵ is to be 5%, 2%, and 1%.

***12.23** A class AB output stage, resembling that in Fig. 12.10 but utilizing a single supply of +10 V and biased at $V_I = 6$ V, is capacitively coupled to a 100-Ω load. For transistors for which $|V_{BE}| = 0.7$ V at 10 mA and for a bias voltage $V_{BB} = 1.3$ V, what quiescent current results? For a step change in output from 0 to −1 V, what input step is required? Assuming transistor-saturation voltages of zero, find the largest possible positive-going and negative-going steps at the output.

Section 12.5: Biasing the Class AB Circuit

D 12.24 Consider the diode-biased class AB circuit of Fig. 12.13. For $I_{BIAS} = 200\,\mu$A, find the relative size (n) that should be used for the output devices (in comparison to the biasing devices) to ensure that an output resistance of 5 Ω or less is obtained in the quiescent state. Neglect the small-signal resistance of the biasing diodes.

D *12.25 A class AB output stage using a two-diode bias network as shown in Fig. 12.13 utilizes diodes having the same junction area as the output transistors. For $V_{CC} = 10$ V, $I_{BIAS} = 2$ mA, $R_L = 50\,\Omega$, $\beta_N = 50$, and $|V_{CEsat}| = 0$ V, what is the quiescent current? What are the largest possible positive and negative output signal levels? To achieve a positive peak output level equal to the negative peak level, what value of β_N is needed if I_{BIAS} is not changed? What value of I_{BIAS} is needed if β_N is held at 50? For this value, what does I_Q become?

D 12.26 Evaluate the small-signal input resistance and small-signal voltage gain of the class AB output stage in Fig. 12.13. To simplify matters, assume the small-signal resistances of D_1 and D_2 to be negligibly small. Replace each of Q_N and Q_P with its hybrid-π model and neglect r_o. Hence show that the class AB stage is equivalent, from a small-signal point of view, to an emitter-follower transistor whose $r_\pi = r_{\pi N} \| r_{\pi P}$ and $g_m = g_{mN} + g_{mP}$, and hence $r_e = r_{eN} \| r_{eP}$ and $\beta = (g_{mN} + g_{mP})(r_{\pi N} \| r_{\pi P})$. Now show that

$$\frac{v_o}{v_i} = \frac{R_L}{R_L + (r_{eN} \| r_{eP})}$$

and

$$R_{in} \simeq \beta[R_L + (r_{eN} \| r_{eP})]$$

12.27 Figure P12.27 shows a class AB output stage with a common-emitter transistor added to increase the voltage gain and reduce the current that v_I has to supply. Neglecting the small-signal resistances of D_1 and D_2, find the small-signal voltage gain v_o/v_i. (*Hint*: Use the expressions for voltage gain and input resistance of the class AB stage without Q_3, given in the statement for Problem 12.26.)

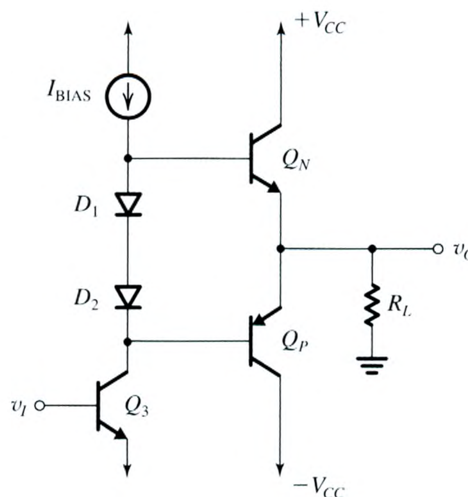


Figure P12.27

12.28 Find an expression for the output resistance R_{out} of the class AB output stage in Fig. P12.27. Toward that end, neglect the small-signal resistance of each of D_1 and D_2 and assume the current source supplying I_{BIAS} has an output resistance R_{BIAS}. Transistors Q_N and Q_P are equivalent to a single transistor with $r_\pi = r_{\pi N} \| r_{\pi P}$, $r_e = r_{eN} \| r_{eP}$, and $g_m = g_{mN} + g_{mP}$.

****12.29** A class AB output stage using a two-diode bias network as shown in Fig. 12.13 utilizes diodes having the same junction area as the output transistors. At a room temperature of about 20°C the quiescent current is 1 mA and $|V_{BE}| = 0.6$ V. Through a manufacturing error, the thermal coupling between the output transistors and the biasing diode-connected transistors is omitted. After some output activity, the output devices heat up to 70°C while the biasing devices remain at 20°C. Thus, while the V_{BE} of each device remains unchanged, the quiescent current in the output devices increases. To calculate the new current value, recall that there are two effects: I_S increases by about 14%/°C and $V_T = kT/q$ changes, where $T = 273° +$ temperature in °C, and $V_T = 25$ mV only at 20°C. However, you may assume that β_N remains almost constant. This assumption is based on the fact that β increases with temperature but decreases with current. What is the new value of I_Q? If the power supply is ±20 V, what additional power is dissipated? If thermal runaway occurs, and the temperature of the output transistors increases by 10°C for every watt of additional power

dissipation, what additional temperature rise and current increase result?

D 12.30 Repeat Example 12.6 for the situation in which the peak positive output current is 200 mA. Use the same general approach to safety margins. What are the values of R_1 and R_2 you have chosen?

****12.31** A V_{BE} multiplier is designed with equal resistances for nominal operation at a terminal current of 1 mA, with half the current flowing in the bias network. The initial design is based on $\beta = \infty$ and $V_{BE} = 0.7$ V at 1 mA.

(a) Find the required resistor values and the terminal voltage.
(b) Find the terminal voltage that results when the terminal current increases to 2 mA. Assume $\beta = \infty$.
(c) Repeat (b) for the case the terminal current becomes 20 mA.
(d) Repeat (c) using the more realistic value of $\beta = 100$.

***12.32** By replacing the transistor in the V_{BE} multiplier by its hybrid-π, small-signal model (with r_o neglected), show that the incremental resistance between the two terminals of the multiplier is given by

$$r = \frac{R_2 + (R_1 \| r_\pi)}{1 + g_m(R_1 \| r_\pi)}$$

Evaluate r for the case $R_1 = R_2 = 1.2$ kΩ, with the transistor operating at $I_C = 1$ mA and having $\beta = 100$.

12.33 Use the results given in the answer to Exercise 12.9 to determine the input current of the circuit in Fig. 12.16 for $v_I = 0$ and ± 10 V with infinite and 100-Ω loads.

12.34 For the circuit in Fig 12.16, operated near $v_I = 0$ and fed with a signal source having zero resistance, show that the output resistance is given by

$$R_{\text{out}} = \frac{1}{2}\left[R_3 + r_{e3} + (R_1 \| r_{e1})/(\beta_3 + 1)\right]$$

Assume that the top and bottom halves of the circuit are perfectly matched.

D ***12.35 Consider the circuit of Fig. 12.16 in which Q_1 and Q_2 are matched, and Q_3 and Q_4 are matched but have three times the junction area of the others. Resistors R_3 and R_4 also are matched. For $V_{CC} = 10$ V, find values for resistors R_1 through R_4 that allow for a base current of at least 10 mA in Q_3 (and Q_4) at $v_I = +5$ V ($v_I = -5$ V), when a load demands it, with at most a 2-to-1 variation in currents in Q_1 (and Q_2). The quiescent current in Q_3 is to be 40 mA. Let $\beta_{1,2} \geq 150$ and $\beta_{3,4} \geq 50$. For input voltages around 0 V, estimate the output resistance of the overall follower driven by a source having zero resistance. For an input voltage of +1 V and a load resistance of 2 Ω, what output voltage results? Q_1 and Q_2 have $|V_{BE}|$ of 0.7 V at a current of 10 mA.

12.36 Figure P12.36 shows a variant of the class AB circuit of Fig. 12.16. Assume that all four transistors are matched with $\beta = 100$ except that the emitter area of $Q_{3,4}$ is 10 times that of $Q_{1,2}$.

(a) For $v_I = 0$, find the quiescent current in Q_3 and Q_4, the input current i_I, and the output voltage v_O.
(b) Since the circuit has perfect symmetry, the small-signal performance around $v_I = 0$ can be determined by considering either the top or bottom half of the circuit only. In this case, the load on the half-circuit must be $2R_L$, the input resistance found is $2R_{\text{in}}$, and the output resistance found is $2R_{\text{out}}$. Using this approach, find R_{in}, v_o/v_i, and R_{out} (assuming that the circuit is fed with a zero-resistance source).

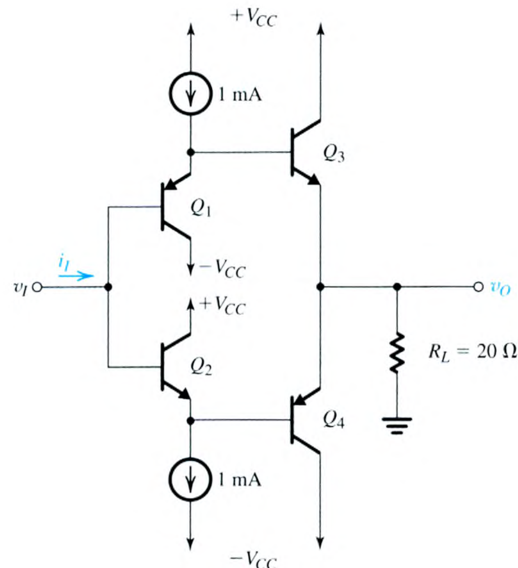


Figure P12.36

12.37 For the Darlington configuration shown in Fig. 12.17, show that for $\beta_1 \gg 1$ and $\beta_2 \gg 1$:

(a) The equivalent composite transistor has $\beta \simeq \beta_1\beta_2$.
(b) If the composite transistor is operated at a current I_C, then Q_2 will be operating at a collector current approximately equal to I_C, and Q_1 will be operating at a collector current approximately equal to I_C/β_2.
(c) The composite transistor has a base–emitter voltage $V_{BE} \simeq 2V_T\ln(I_C/I_S) - V_T\ln(\beta_2)$, where I_S is the saturation current of each of Q_1 and Q_2.
(d) The composite transistor has an equivalent $r_\pi \simeq 2\beta_1\beta_2(V_T/I_C)$.
(e) The composite transistor has an equivalent $g_m \simeq \frac{1}{2}(I_C/V_T)$.

***12.38** For the circuit in Fig. P12.38 in which the transistors have $V_{BE} = 0.7$ V and $\beta = 100$:

(a) Find the dc collector current for each of Q_1 and Q_2.
(b) Find the small-signal current i_c that results from an input signal v_i, and hence find the voltage gain v_o/v_i.
(c) Find the input resistance R_{in}.

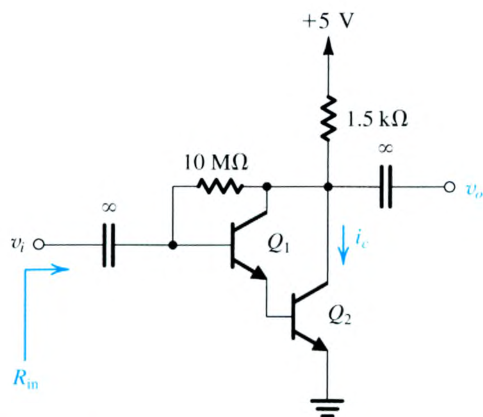


Figure P12.38

SIM **12.39** The BJTs in the circuit of Fig. P12.39 have $\beta_P = 10$, $\beta_N = 100$, $|V_{BE}| = 0.7$ V, and $|V_A| = 100$ V.

(a) Find the dc collector current of each transistor and the value of V_C.
(b) Replacing each BJT with its hybrid-π model, show that

$$\frac{v_o}{v_i} \simeq g_{m1}\left[r_{o1} \,\|\, \beta_N(r_{o2} \,\|\, R_f)\right]$$

(c) Find the values of v_o/v_i and R_{in}.

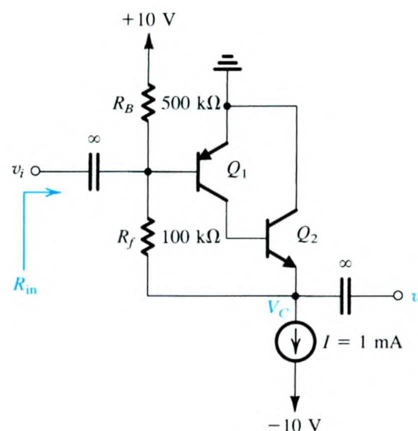


Figure P12.39

D **12.40 Consider the compound-transistor class AB output stage shown in Fig. 12.19 in which Q_2 and Q_4 are matched transistors with $V_{BE} = 0.7$ V at 10 mA and $\beta = 100$, Q_1 and Q_5 have $V_{BE} = 0.7$ V at 1-mA currents and $\beta = 100$, and Q_3 has $V_{EB} = 0.7$ V at a 1-mA current and $\beta = 10$. Design the circuit for a quiescent current of 1 mA in Q_2 and Q_4, I_{BIAS} that is 100 times the standby base current in Q_1, and a current in Q_5 that is nine times that in the associated resistors. Find the values of the input voltage required to produce outputs of ± 10 V for a 1-kΩ load. Use V_{CC} of 15 V.

***12.41** Figure P12.41 shows a variant on the class AB amplifier known as class G. Here, in addition to the normal power supply $\pm V_{CC1}$, the circuit is equipped with a higher voltage supply $\pm V_{CC2}$. The latter supply is utilized only infrequently. The circuit operates as follows. Normally, D_1 and D_2 are turned on and thus connect the $\pm V_{CC1}$ supply to the class AB stage transistors Q_1 and Q_2. Simultaneously, Q_3 and Q_4 are off. For v_I positive and exceeding a certain threshold, Q_3 turns on, D_1 turns off, and Q_1 is then effectively operating from the higher voltage supply V_{CC2}. This continues as long as v_I is larger than the specified threshold. As v_I decreases below the threshold value, Q_3 is turned off and D_1 turns on, thus connecting Q_1 to its normal supply V_{CC1}. A similar process happens in the negative direction, with D_2 and Q_4 taking the place of D_1 and Q_3. Let $V_{CC1} = 35$ V, $V_{CC2} = 70$ V, $V_{Z1} = 3.3$ V, and the voltage of the V_{BE} multiplier $V_{BB} = 1.2$ V.

(a) Find the positive threshold value of v_I at which Q_3 is turned on.
(b) If for 95% of the time v_I is in the vicinity of 30 V and only 5% of the time it is in the vicinity of 65 V, use Eq. (12.19) to estimate the average power dissipated in the

transistors, P_D. Compare to the value of P_D dissipated in a class AB stage operated from a ±70 V supply.

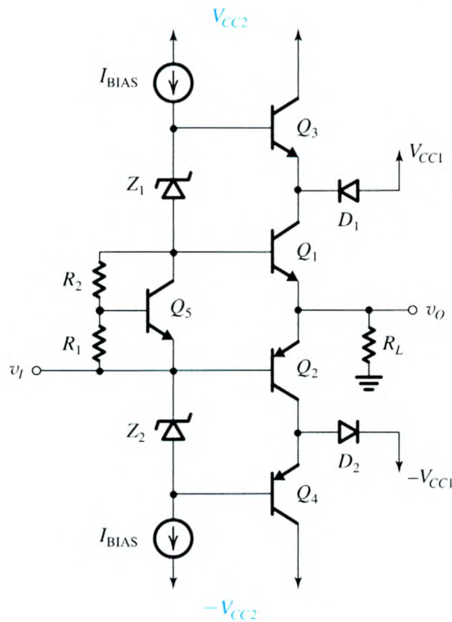


Figure P12.41

12.42 Figure P12.42 shows a class AB output stage equipped with protection against the effect of short-circuiting the output while the stage is sourcing current. Find the value of R_{E1} that causes Q_5 to turn on and absorb all of the 2-mA bias current when the output current being sourced reaches 100 mA, robbing Q_1 of its base current. For Q_5, $I_S = 10^{-14}$ A. If the normal peak output current is 60 mA, find the voltage drop across R_{E1} and the collector current of Q_5.

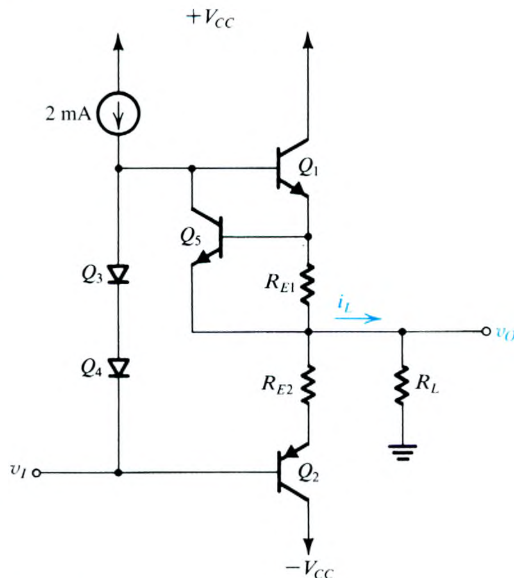


Figure P12.42

D 12.43 The circuit shown in Fig. P12.43 limits the output current from Q_3 in the event of a short circuit or other mishap. It has the advantage that the current-sensing resistor R does not appear directly at the output. Find the value of R that causes Q_5 to turn on and absorb all of $I_{BIAS} = 2$ mA, when the current being sourced reaches 150 mA. For Q_5, $I_S = 10^{-14}$ A. If the normal peak output current is 75 mA, find the voltage drop across R and the collector current in Q_5.

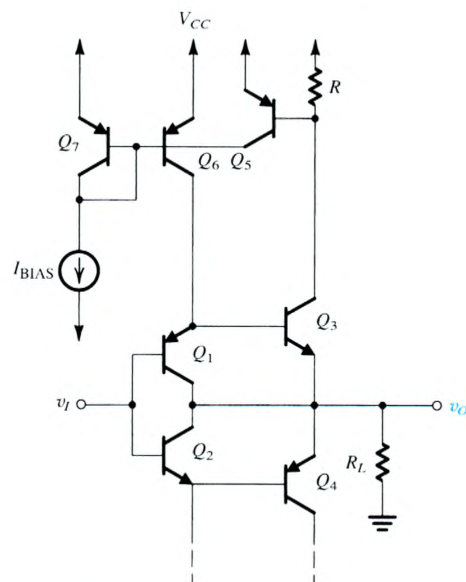


Figure P12.43

Section 12.6: CMOS Output Stages

12.44 For the source follower in Fig. 12.20, assume $V_{DD} = V_{SS} = 2.5$ V with all transistors having $k_n = 50$ mA/V^2 and $V_{tn} = 0.5$ V. If $I = 15$ mA and $R_L = 150\ \Omega$, find the drain current i_{D1}, input voltage v_I, and incremental gain corresponding to $v_O = +1$ V, 0 V, and -1 V.

12.45 What bias current I is required in Fig. 12.20 to ensure that the output resistance is only 5 Ω with no load current,

assuming $V_{OV1} = 350$ mV? Compare this to the current I required to achieve the same output resistance for the emitter follower in Fig. 12.3.

12.46 The class AB output stage in Fig. 12.23 utilizes two matched transistors with $k_n = k_p = 250$ mA/V^2 and is operated from ±2.5-V power supplies. If the stage is required to supply a maximum current of ±20 mA while keeping Q_N and Q_P in saturation, what is the output voltage swing realized?

12.47 For the CMOS output stage of Fig. 12.23 with $I_Q =$ 1 mA, $|V_{OV}| = 0.2$ V for each of Q_P and Q_N at the quiescent point, and $\mu = 5$, find the output resistance at the quiescent point.

12.48 (a) Show that for the CMOS output stage of Fig. 12.23, assuming $R_{out} \ll R_L$, the deviation of the gain from unity is

$$|\text{Gain error}| = \frac{R_{out}}{R_L}$$

(b) For a stage that drives a load resistance of 200 Ω with a gain error of less than 3%, find the overdrive voltage at which Q_P and Q_N should be operated. Let $I_Q = 2$ mA and $\mu = 5$.

12.49 Show that in the CMOS class AB common-source output stage (Fig. 12.23), Q_N turns off when $v_O = 4I_Q R_L$ and Q_P turns off when $v_O = -4I_Q R_L$. This is equivalent to saying that one of the transistors turns off when $|i_L|$ reaches $4I_Q$.

D *12.50 Design the circuit of Fig. 12.23 to drive a load resistance of 50 Ω while exhibiting an output resistance, around the quiescent point, of 2.5 Ω. Operate Q_N and Q_P at $I_Q = 1.5$ mA and $|V_{OV}| = 0.15$ V. The technology utilized is specified to have $k_n' = 250$ μA/V^2, $k_p' = 100$ μA/V^2, $V_{tn} = -V_{tp} = 0.5$ V, and $V_{DD} = V_{SS} = 2.5$ V.

(a) Specify (W/L) for each of Q_N and Q_P.
(b) Specify the required value of μ.
(c) What is the expected error in the stage gain?
(d) In the quiescent state, what dc voltage must appear at the output of each of the error amplifiers?
(e) At what value of positive v_O will Q_P be supplying all the load current? Repeat for negative v_O and Q_N supplying all the load current.
(f) What is the linear range of v_O?

12.51 Sketch waveforms resembling those in Fig. 12.24(a). Let v_T have ±10 V peaks and assume v_A is a sine wave with 5-V peak amplitude. Let the frequency of v_T be 5 times that of v_A. The comparator output levels are ±10 V.

12.52 A pulse waveform swinging between ±10 V has a duty ratio of 0.7. What is its average value? If the duty ratio is changed to 0.3, what does the average value become?

12.53 Consider the circuit in Fig. 12.24(b). If v_A is a sine wave, what is the maximum power supplied to a load of resistance R, in terms of V_{DD}?

Section 12.7 Power Transistors

12.54 The emitter follower in Fig. 12.3 drives a load resistance $R_L = 50$ Ω with an output ranging between $v_O =$ ±5 V. It operates with $V_{CC} = 10$ V and $I = 200$ mA. What is the peak power dissipation in Q_1 and what is the corresponding output voltage v_O? What is the peak power dissipation in Q_2 and what is the corresponding v_O? If identical transistors are to be used for Q_1 and Q_2, what minimum specifications are required for $I_{C\max}$ and BV_{CEO} allowing a factor-of-2 safety margin?

12.55 A power MOSFET is specified to have $I_{D\max} = 5$ A, $V_{DS\max} = 50$ V, and $P_{D\max} = 50$ W.

(a) Sketch the SOA boundaries.
(b) If the MOSFET is used in the common-source configuration as shown in Fig. P12.55, show that the maximum current occurs when $V_{DS} = 0$, the maximum V_{DS} occurs when $I_D = 0$, and the maximum power dissipation occurs when $V_{DS} = V_{DD}/2$.
(c) For $V_{DD} = 40$ V, find the smallest resistance R for which the operating point is always within the SOA. What are the corresponding values of $I_{D\max}$ and $P_{D\max}$?
(d) Repeat (c) for $V_{DD} = 30$ V.
(e) Repeat (c) for $V_{DD} = 15$ V.

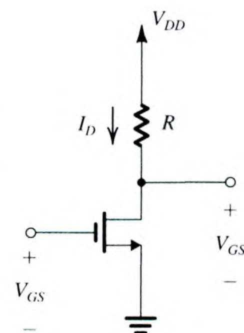


Figure P12.55

CHAPTER 13

Operational-Amplifier Circuits

IN THIS CHAPTER YOU WILL LEARN

- The design and analysis of the two basic CMOS op-amp architectures: the two-stage circuit and the single-stage, folded-cascode circuit.
- Interesting and useful circuits for the design of BJT op amps.
- How to break a large analog circuit into its recognizable blocks to be able to make the analysis amenable to a pencil-and-paper approach, which is the best way to learn design.
- Some of the modern techniques used in the design of low-voltage op amps.
- Most important, how the different topics we studied in the preceding chapters come together in the design of the most important analog IC, the op amp.

Introduction

In this chapter, we will study the internal circuitry of the most important analog IC, the operational amplifier. We covered the terminal characteristics and some circuit applications of op amps in Chapter 2. Here, our objective is to survey some of the ingenious techniques that have evolved for combining elementary analog circuit building blocks to realize a complete op amp. We shall study both CMOS and bipolar op amps.

The CMOS op-amp circuits we will consider are mainly used in the design of analog and mixed-signal VLSI circuits. Because these op amps are usually designed with a specific application in mind, they can be optimized to relax certain specifications. For instance, many CMOS op amps are utilized within an IC and do not connect to the outside terminals of the chip. As a result, the loads on their outputs are usually limited to small capacitances of at most a few picofarads. Internal CMOS op amps therefore do not need to have low output resistances, and their design rarely incorporates an output stage. Also, if the op-amp input terminals are not connected to the chip terminals, there will be no danger of static charge damaging the gate oxide of the input MOSFETs. Hence, internal CMOS op amps do not need input clamping diodes for gate protection and thus do not suffer from the leakage effects of such diodes. In other words, they may take full advantage of near-infinite input resistance of the MOSFET.

While CMOS op amps are used extensively in the design of VLSI systems, the BJT remains widely used in the design of general-purpose op amps. These op amps are used in a wide variety of applications and are designed to fit a wide range of specifications. As a result, the circuit of a general-purpose op amp represents a compromise among many performance parameters. We will study modern design techniques for general-pupose BJT op amps in the last section.

In addition to presenting some of the ideas that make analog IC design such an exciting topic, this chapter should tie together many of the concepts and methods we have studied so far.

THE GENIE OF ANALOG

The need for precision in analog ICs supported a generation of highly skilled circuit and process engineers whose creative approaches to overcoming technology limitations made them celebrities among their peers. Perhaps most famous of all was Robert Widlar, who teamed with process engineer Dave Talbert to bootstrap the analog business in the 1960s with highly successful designs for Fairchild and National Semiconductor. Widlar's pranks, including threatening to cut through bureaucracy with an axe and buying a sheep to trim National's unkempt lawns, remain Silicon Valley legends.

Robert John Widlar was already a legendary chip designer at age 33, and a pioneer of linear analog integrated-circuit design. He created the Widlar current source, the Widlar bandgap voltage reference, the Widlar output stage, and a host of op-amp designs, including the first mass-produced operational amplifier ICs (Fairchild μA702, μA709), the first integrated voltage regulator (μA723, National LM100), the first fully internally compensated operational amplifier (LM101), the field-effect input (LM101A), and the super-beta input (LM108). Each of Widlar's designs became a product champion in its class.

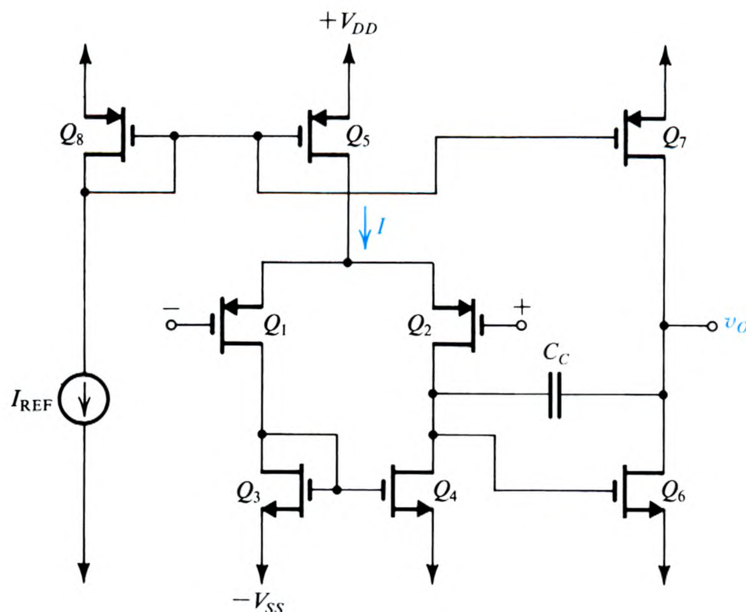


Figure 13.1 The basic two-stage CMOS op-amp configuration.

13.1 The Two-Stage CMOS Op Amp

The first op-amp circuit we study is the two-stage CMOS topology shown in Fig. 13.1. This simple but elegant circuit has become a classic and is used in a variety of forms in the design of VLSI systems. We studied this circuit in Section 9.6.1 as an example of a multistage CMOS amplifier and urge the reader to review that material before proceeding. Here, our detailed study will emphasize the performance characteristics of the circuit and the trade-offs involved in its design.

13.1.1 The Circuit

The circuit consists of two gain stages: The first is formed by the differential pair Q_1–Q_2 together with its current-mirror load Q_3–Q_4. This differential-amplifier circuit, studied in detail in Section 9.5, provides a voltage gain that may range from 10 V/V to 60 V/V, as well as converting from differential to single-ended form while providing common-mode rejection.

The differential pair is biased by current source Q_5, which is one of the two output transistors of the current mirror formed by Q_8, Q_5, and Q_7. The current mirror is fed by a reference current I_{REF}, which can be generated by simply connecting a resistor to the negative supply voltage $-V_{SS}$. Circuits that generate a more precise I_{REF} exist for applications with more stringent requirements.

The second gain stage consists of the common-source transistor Q_6 and its current-source load Q_7. The second stage typically provides a gain of 8 V/V to 80 V/V. In addition, it takes part in the process of frequency compensating the op amp. From Section 11.10, you will recall that to guarantee that the op amp will operate in a stable fashion (as opposed to oscillating) when negative feedback of various amounts is applied, the open-loop gain is

made to roll off with frequency at the uniform rate of –20 dB/decade. This in turn is achieved by introducing a pole at a relatively low frequency and arranging for it to be dominant in determining the frequency response. In the circuit we are studying, we implement this using a compensation capacitance C_C connected in the negative-feedback path of the second-stage amplifying transistor Q_6. As we will see, C_C (together with the much smaller capacitance C_{gd6} across it) is Miller-multiplied by the gain of the second stage, and the resulting capacitance at the input of the second stage interacts with the output resistance of the first stage to provide the required dominant pole (more on this later).

Unless properly designed, the CMOS op-amp circuit in Fig. 13.1 can exhibit a **systematic output dc offset** voltage. In Section 9.6.1, we found that the systematic dc offset can be eliminated by sizing the transistors to satisfy the following constraint:

$$\frac{(W/L)_6}{(W/L)_4} = 2\frac{(W/L)_7}{(W/L)_5} \tag{13.1}$$

Finally, notice that the CMOS op-amp circuit of Fig. 13.1 does not have an output stage. This is because it is usually required to drive only small on-chip capacitive loads.[1]

13.1.2 Input Common-Mode Range and Output Swing

Refer to Fig. 13.1 and consider the situation when the two input terminals are tied together and connected to a voltage V_{ICM}. The lowest value of V_{ICM} has to be large enough to keep Q_1 and Q_2 in saturation. Thus, the lowest value of V_{ICM} should not be lower than the voltage at the drain of Q_1 $(-V_{SS} + V_{GS3} = -V_{SS} + V_{tn} + V_{OV3})$ by more than $|V_{tp}|$, thus

$$V_{ICM} \geq -V_{SS} + V_{tn} + V_{OV3} - |V_{tp}| \tag{13.2}$$

The highest value of V_{ICM} should ensure that Q_5 remains in saturation; that is, the voltage across Q_5, V_{SD5}, should not decrease below $|V_{OV5}|$. Equivalently, the voltage at the drain of Q_5 should not go higher than $V_{DD} - |V_{OV5}|$. Thus the upper limit of V_{ICM} is

$$V_{ICM} \leq V_{DD} - |V_{OV5}| - V_{SG1}$$

or equivalently

$$V_{ICM} \leq V_{DD} - |V_{OV5}| - |V_{tp}| - |V_{OV1}| \tag{13.3}$$

We can combine Eqs. (13.2) and (13.3) to express the input common-mode range as

$$-V_{SS} + V_{OV3} + V_{tn} - |V_{tp}| \leq V_{ICM} \leq V_{DD} - |V_{tp}| - |V_{OV1}| - |V_{OV5}| \tag{13.4}$$

As expected, the overdrive voltages, which are important design parameters, subtract from the dc supply voltages, thereby reducing the input common-mode range. From a V_{ICM} range point of view, we want to select the values of V_{OV} as low as possible. We can see from Eq. (13.4) that the lower limit of V_{ICM} is approximately within an overdrive voltage of $-V_{SS}$. The upper limit, however, is not as good; it is lower than V_{DD} by two overdrive voltages and a threshold voltage.

---

[1] If the amplifier must drive low-resistance loads and thus needs a low output resistance, we can connect a source follower or a more complex output stage from Chapter 12 to the output of the second stage.

The extent of the signal swing allowed at the output of the op amp is limited at the lower end by the need to keep Q_6 saturated and at the upper end by the need to keep Q_7 saturated, thus

$$-V_{SS} + V_{OV6} \leq v_O \leq V_{DD} - |V_{OV7}| \tag{13.5}$$

Thus the output voltage can swing to within an overdrive voltage of each of the supply rails. This is a reasonably wide output swing and can be maximized by selecting values for $|V_{OV}|$ of Q_6 and Q_7 as low as possible.

If we wish to use an op amp as a unity-gain feedback amplifier, it must be possible to connect its output terminal to its negative input terminal. For such a connection to be possible, there must be a substantial overlap between the allowable range of v_O and the allowable range of V_{ICM}. This is usually the case in the basic two-stage CMOS op-amp circuit.

EXERCISE

13.1 For a particular design of the two-stage CMOS op amp of Fig. 13.1, ±1.65-V supplies are used and all transistors except for Q_6 and Q_7 are operated with overdrive voltages of 0.3-V magnitude; Q_6 and Q_7 use overdrive voltages of 0.5-V magnitude, and $V_{tn} = |V_{tp}| = 0.5$ V. Find the input common-mode range and the range allowed for v_O. If the op amp is used in a unity-gain configuration, what is the allowable range of output voltage?
Ans. −1.35 V to 0.55 V; −1.15 V to +1.15 V; −1.15 V to +0.55 V

13.1.3 DC Voltage Gain

To determine the dc voltage gain and the frequency response, consider a simplified equivalent-circuit model for the small-signal operation of the CMOS two-stage op amp (Fig. 13.2), where each of the two stages is modeled as a transconductance amplifier. As expected, the input resistance is practically infinite,

$$R_{in} = \infty$$

The first-stage transconductance G_{m1} is equal to the transconductance of each of Q_1 and Q_2 (see Section 9.5),

$$G_{m1} = g_{m1} = g_{m2} \tag{13.6}$$

Since Q_1 and Q_2 operate at equal bias currents ($I/2$) and equal overdrive voltages, $|V_{OV1}| = |V_{OV2}|$,

$$G_{m1} = \frac{2(I/2)}{|V_{OV1}|} = \frac{I}{|V_{OV1}|} \tag{13.7}$$

Resistance R_1 represents the output resistance of the first stage, thus

$$R_1 = r_{o2} \| r_{o4} \tag{13.8}$$

where

$$r_{o2} = \frac{|V_{A2}|}{I/2} \tag{13.9}$$

and

$$r_{o4} = \frac{V_{A4}}{I/2} \tag{13.10}$$

The dc gain of the first stage is thus

$$A_1 = -G_{m1}R_1 = -g_{m1}(r_{o2} \| r_{o4}) = -\frac{2}{|V_{OV1}|} \Big/ \left[\frac{1}{|V_{A2}|} + \frac{1}{V_{A4}}\right] \tag{13.11}$$

Observe that the magnitude of A_1 is increased by operating the differential-pair transistors, Q_1 and Q_2, at a low overdrive voltage, and by choosing longer channel lengths for Q_1, Q_2, Q_3, and Q_4 so as to obtain larger Early voltages, $|V_A|$.

Returning to the equivalent circuit in Fig. 13.2 and leaving the discussion of the various model capacitances until Section 13.1.5, we note that the second-stage transconductance G_{m2} is given by

$$G_{m2} = g_{m6} = \frac{2I_{D6}}{V_{OV6}} \tag{13.12}$$

Resistance R_2 represents the output resistance of the second stage, thus

$$R_2 = r_{o6} \| r_{o7} \tag{13.13}$$

where

$$r_{o6} = \frac{V_{A6}}{I_{D6}} \tag{13.14}$$

and

$$r_{o7} = \frac{|V_{A7}|}{I_{D7}} = \frac{|V_{A7}|}{I_{D6}} \tag{13.15}$$

The voltage gain of the second stage can now be found as

$$A_2 = -G_{m2}R_2 = -g_{m6}(r_{o6} \| r_{o7}) = -\frac{2}{V_{OV6}} \Big/ \left[\frac{1}{V_{A6}} + \frac{1}{|V_{A7}|}\right] \tag{13.16}$$

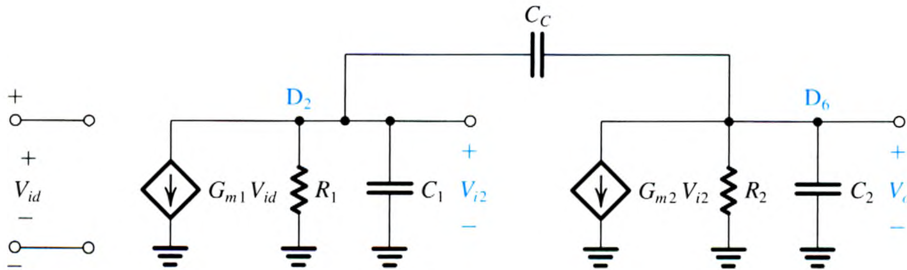


Figure 13.2 Small-signal equivalent circuit for the op amp in Fig. 13.1.

Here again we observe that to increase the magnitude of A_2, we operate Q_6 at a low overdrive voltage and use longer channel lengths for Q_6 and Q_7.

The overall dc voltage gain is the product A_1A_2,

$$A_v = A_1A_2 = G_{m1}R_1G_{m2}R_2 = g_{m1}(r_{o2} \| r_{o4})g_{m6}(r_{o6} \| r_{o7}) \tag{13.17}$$

Note that A_v is of the order of $(g_mr_o)^2$. Thus the value of A_v will be in the range of 100 V/V to 5000 V/V. It is largely determined by the intrinsic gain of transistors $Q_{1,2}$ $(g_{m1,2}r_{o1,2})$ and Q_6 $(g_{m6}r_{o6})$, albeit with some contribution from r_{o4} and r_{o7}.

Finally, we note that the output resistance of the op amp is equal to the output resistance of the second stage,

$$R_o = r_{o6} \| r_{o7} \tag{13.18}$$

Hence R_o can be large (i.e., in the tens-of-kilohms range). Nevertheless, as we learned from Chapter 11, negative feedback that samples the op-amp output voltage reduces the output resistance by a factor equal to the amount of feedback $(1 + A\beta)$. Also, as mentioned before, CMOS op amps within integrated circuits rarely need to drive heavy resistive loads. Smaller values of R_o (i.e., of the order of a kilohm) are realized with very short gate lengths (e.g., $L_{6,7} < 0.5\,\mu$m) and large bias current I_{D6}.

EXERCISES

13.2 The CMOS op amp of Fig. 13.1 is fabricated in the 0.13-μm process of Appendix K. Find A_1, A_2, and A_v if all devices are 0.5 μm long, $|V_{OV1}| = 0.2$ V, and $V_{OV6} = 0.3$ V. Also, find the op-amp output resistance obtained when the second stage is biased at 0.1 mA.
Ans. −13.6 V/V; –9.1 V/V; 124 V/V; 13.6 kΩ

13.3 If the CMOS op amp in Fig. 13.1 is connected as a unity-gain buffer, show that the closed-loop output resistance is given by $R_{\text{out}} \simeq 1/g_{m6}\left[g_{m1}\left(r_{o2} \| r_{o4}\right)\right]$.

13.1.4 Common-Mode Rejection Ratio (CMRR)

The CMRR of the two-stage op amp of Fig. 13.1 is determined by the first stage. We analyzed this in Section 9.5.5 and the result is given in Eq. (9.138), namely,

$$\text{CMRR} = \left[g_{m1}(r_{o2} \| r_{o4})\right]\left[2g_{m3}R_{SS}\right] \tag{13.19}$$

where R_{SS} is the output resistance of the bias current source $Q_5(r_{o5})$. Observe that CMRR is of the order of $(g_mr_o)^2$ and thus can be reasonably high. Also, since g_mr_o is proportional to $V_A/V_{OV} = V_A'L/V_{OV}$, the CMRR is increased if we use long channels, especially for Q_5, and low overdrive voltages for Q_{1-4}.

13.1.5 Frequency Response

Refer to the equivalent circuit in Fig. 13.2. Capacitance C_1 models all the capacitance between the output node of the first stage and ground, thus may be approximated as[2]

$$C_1 = C_{gd2} + C_{db2} + C_{gd4} + C_{db4} + C_{gs6} \tag{13.20}$$

Capacitance C_2 represents the total capacitance between the output node of the op amp and ground and includes whatever load capacitance C_L that the amplifier is required to drive, thus approximately[3]

$$C_2 = C_{db6} + C_{db7} + C_{gd7} + C_L \tag{13.21}$$

Usually, C_L is larger than the transistor capacitances, so C_2 becomes much larger than C_1. As we mentioned before, we deliberately included capacitor C_C to equip the op amp with a uniform –6-dB/octave frequency response. In the following, we see how this is possible and how to select a value for C_C. Finally, note that in the equivalent circuit of Fig. 13.2 we should have included C_{gd6} in parallel with C_C. Usually, however, $C_C \gg C_{gd6}$, which is why we have neglected C_{gd6}.

To determine V_o, we analyze the circuit in Fig. 13.2 as follows. Writing a node equation at node D_2 yields

$$G_{m1}V_{id} + \frac{V_{i2}}{R_1} + sC_1V_{i2} + sC_C(V_{i2} - V_o) = 0 \tag{13.22}$$

Writing a node equation at node D_6 yields

$$G_{m2}V_{i2} + \frac{V_o}{R_2} + sC_2V_o + sC_C(V_o - V_{i2}) = 0 \tag{13.23}$$

To eliminate V_{i2} and thus determine V_o in terms of V_{id}, we use Eq. (13.23) to express V_{i2} in terms of V_o and substitute the result into Eq. (13.22). After some straightforward manipulations we obtain the amplifier transfer function

$$\frac{V_o}{V_{id}} = \frac{G_{m1}(G_{m2} - sC_C)R_1R_2}{1 + s[C_1R_1 + C_2R_2 + C_C(G_{m2}R_1R_2 + R_1 + R_2)] + s^2[C_1C_2 + C_C(C_1 + C_2)]R_1R_2} \tag{13.24}$$

First we note that for $s = 0$ (i.e., dc), Eq. (13.24) gives $V_o/V_{id} = (G_{m1}R_1)(G_{m2}R_2)$, which is what we should have expected. Second, the transfer function in Eq. (13.24) indicates that the amplifier has a transmission zero determined by solving $G_{m2} - s_ZC_C = 0$ for s_Z. Thus,

$$s_Z = \frac{G_{m2}}{C_C} \tag{13.25}$$

In other words, the zero is on the positive real axis with a frequency ω_Z of

$$\omega_Z = \frac{G_{m2}}{C_C} \tag{13.26}$$

[2] The gates of Q_2 and Q_4 are not true small-signal ground, so C_{gd2} and C_{gd4} are not precisely captured in Fig. 13.2.

[3] The gate of Q_7 is not a small-signal ground due to the small-signal resistance of diode-connected Q_8, so C_{gd7} is not precisely captured by Fig. 13.2.

Also, the amplifier has two poles that are the roots of the denominator polynomial of Eq. (13.24). If the frequencies of the two poles are denoted ω_{P1} and ω_{P2}, then the denominator polynomial can be expressed as

$$D(s) = \left(1 + \frac{s}{\omega_{P1}}\right)\left(1 + \frac{s}{\omega_{P2}}\right) = 1 + s\left(\frac{1}{\omega_{P1}} + \frac{1}{\omega_{P2}}\right) + \frac{s^2}{\omega_{P1}\omega_{P2}}$$

Now if one of the poles is dominant, say with frequency ω_{P1}, then $\omega_{P1} \ll \omega_{P2}$, and we can approximate $D(s)$ by

$$D(s) \simeq 1 + \frac{s}{\omega_{P1}} + \frac{s^2}{\omega_{P1}\omega_{P2}} \tag{13.27}$$

We can now find the frequency of the dominant pole, ω_{P1}, by equating the coefficients of the s terms in the denominator in Eq. (13.24) and in Eq. (13.27),

$$\begin{aligned}\omega_{P1} &= \frac{1}{C_1R_1 + C_2R_2 + C_C(G_{m2}R_2R_1 + R_1 + R_2)} \\ &= \frac{1}{R_1\left[C_1 + C_C(1 + G_{m2}R_2)\right] + R_2(C_2 + C_C)}\end{aligned} \tag{13.28}$$

We recognize the first term in the denominator as arising at the interface between the first and second stages. Here, R_1, the output resistance of the first stage, is interacting with the total capacitance at the interface. The latter is the sum of C_1 and the Miller capacitance $C_C(1 + G_{m2}R_2)$, which results from connecting C_C in the negative-feedback path of the second stage whose gain is $G_{m2}R_2$. Now, since R_1 and R_2 are usually of comparable value, we see that the first term in the denominator will be much larger than the second and we can approximate ω_{P1} as

$$\omega_{P1} \simeq \frac{1}{R_1\left[C_1 + C_C(1 + G_{m2}R_2)\right]}$$

A further approximation is possible because C_1 is usually much smaller than the Miller capacitance and $G_{m2}R_2 \gg 1$, thus

$$\omega_{P1} \simeq \frac{1}{R_1C_CG_{m2}R_2} \tag{13.29}$$

We can find the frequency of the second, nondominant pole by equating the coefficients of the s^2 terms in the denominator of Eq. (13.24) and in Eq. (13.27) and substituting for ω_{P1} from Eq. (13.29). The result is

$$\omega_{P2} = \frac{G_{m2}C_C}{C_1C_2 + C_C(C_1 + C_2)}$$

Since $C_1 \ll C_2$ and $C_1 \ll C_C$, we can approximate ω_{P2} as

$$\omega_{P2} \simeq \frac{G_{m2}}{C_2} \tag{13.30}$$

Capacitor C_C has the pole-splitting effect described in Section 11.10.3.

To ensure its stability in a unity-gain configuration, the op amp must have a uniform gain rolloff of –20 dB/decade down to 0 dB. We therefore select the value of the compensation capacitor C_C so that the resulting value of ω_{P1} (Eq. 13.29), when multiplied by the dc gain

$(G_{m1}R_1G_{m2}R_2)$, results in a unity-gain frequency ω_t lower than ω_Z and ω_{P2}. Specifically

$$\omega_t = (G_{m1}R_1G_{m2}R_2)\omega_{P1} = \frac{G_{m1}}{C_C} \tag{13.31}$$

which must be lower than $\omega_Z = \dfrac{G_{m2}}{C_C}$ and $\omega_{P2} \simeq \dfrac{G_{m2}}{C_2}$. Thus, the design must satisfy the following two conditions:

$$\frac{G_{m1}}{C_C} < \frac{G_{m2}}{C_2} \tag{13.32}$$

$$G_{m1} < G_{m2} \tag{13.33}$$

Internally compensated two-stage op amps have a fixed capacitance C_C integrated on-chip. If the op-amp load capacitance increases, so does C_2, eventually exceeding the limit in Eq. (13.32). Thus, op-amp datasheets often specify the maximum load capacitance that may be stably driven in a unity-gain configuration. Some op amps provide pins allowing an external compensation capacitance to be connected in parallel with C_C, increasing its effective value and maintaining Eq. (13.32). However, doing so reduces the unity-gain frequency, Eq. (13.31).

EXERCISE

D13.4 Consider the frequency response of the op amp we analyzed in Chapter 9 (see Example 9.6). Let $C_1 = 0.1$ pF and $C_2 = 2$ pF. Find the value of C_C that results in $f_t = 20$ MHz and verify that f_t is lower than f_Z and f_{P2}. Recall from the results of Example 9.6 that $G_{m1} = 0.35$ mA/V and $G_{m2} = 1$ mA/V.
Ans. $C_C = 2.8$ pF; $f_Z = 57$ MHz; $f_{P2} = 80$ MHz

Simplified Equivalent Circuit The uniform –20-dB/decade gain rolloff obtained at frequencies $f \gg f_{P1}$ but lower than f_{P2} and f_Z suggests that at these frequencies, we can represent the op amp by the simplified equivalent circuit shown in Fig. 13.3. Notice that this attractive simplification is based on the assumption that the gain of the second stage, $|A_2|$, is large, and hence a virtual ground appears at the input terminal of the second stage. The second stage then effectively acts as an integrator that is fed with the output current signal of the first stage; $G_{m1}V_{id}$. Although we derived it for the CMOS amplifier, this simplified equivalent circuit is general and applies to a variety of two-stage op amps, including many bipolar op amps.

Phase Margin Figure 13.4 shows a representative Bode plot for the two-stage op amp. It is also a Bode plot of the loop gain ($A\beta$) when the op amp is in a unity-gain feedback configuration, so we use it for a stability study. Note that at the unity-gain frequency f_t, the phase lag exceeds the 90° caused by the dominant pole at f_{P1}. This so-called excess phase

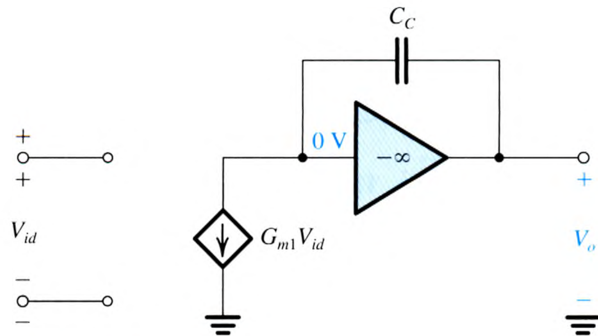


Figure 13.3 An approximate high-frequency equivalent circuit of the two-stage op amp. This circuit applies for frequencies $f \gg f_{P1}$ but lower than f_{P2} and f_Z.

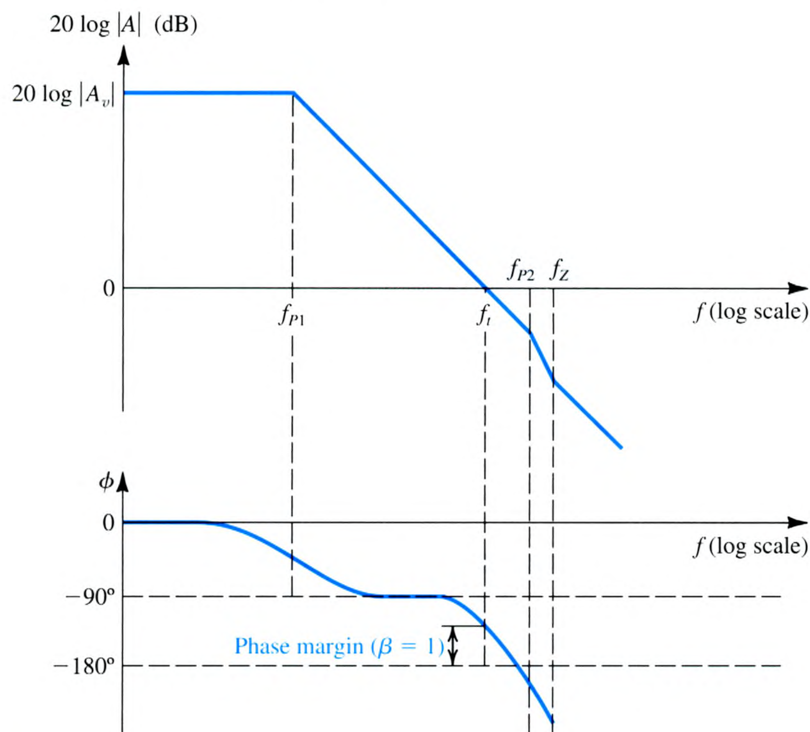


Figure 13.4 Typical frequency response of the two-stage op amp.

shift is due to the second pole,

$$\phi_{P2} = -\tan^{-1}\left(\frac{f_t}{f_{P2}}\right) \tag{13.34}$$

and the right-half-plane zero,

$$\phi_Z = -\tan^{-1}\left(\frac{f_t}{f_Z}\right) \tag{13.35}$$

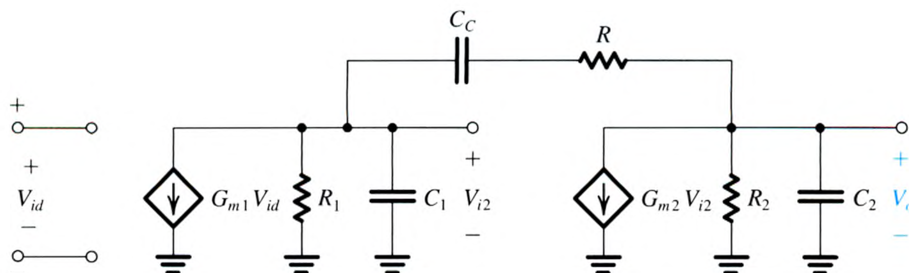


Figure 13.5 Small-signal equivalent circuit of the op amp in Fig. 13.1 with a resistance R included in series with C_C.

Thus the phase lag at $f = f_t$ will be

$$\phi_{\text{total}} = 90° + \tan^{-1}(f_t/f_{P2}) + \tan^{-1}(f_t/f_Z) \tag{13.36}$$

and the phase margin in a unity-gain configuration will be

$$\begin{aligned}\text{Phase margin} &= 180° - \phi_{\text{total}} \\ &= 90° - \tan^{-1}(f_t/f_{P2}) - \tan^{-1}(f_t/f_Z)\end{aligned} \tag{13.37}$$

From our study of the stability of feedback amplifiers in Section 11.9.2, we know that the phase margin significantly affects the closed-loop gain.[4] Therefore, obtaining a desired minimum value of phase margin is usually a design requirement.

The problem of the additional phase lag provided by the right-half-plane zero has a rather simple and elegant solution: By including a resistance R in series with C_C, as shown in Fig. 13.5, we can move the transmission zero to other less harmful locations. To find the new location of the transmission zero, set $V_o = 0$. Then, the current through C_C and R will be $V_{i2}/(R + 1/sC_C)$, and a node equation at the output yields

$$\frac{V_{i2}}{R + \dfrac{1}{sC_C}} = G_{m2}V_{i2} \tag{13.38}$$

Thus the zero is now at

$$s = 1 \Big/ C_C\left(\frac{1}{G_{m2}} - R\right) \tag{13.39}$$

By selecting $R = 1/G_{m2}$, we can place the zero at infinite frequency. An even better choice is often to select R greater than $1/G_{m2}$, thus placing the zero at a negative real-axis location where the phase it introduces becomes a phase lead and thus *adds* to the phase margin.

[4]The amount of phase margin also affects the step response of the closed-loop amplifier.

EXERCISE

13.5 A particular implementation of the CMOS amplifier of Figs. 13.1 and 13.2 provides $G_{m1} = 1$ mA/V, $G_{m2} = 2$ mA/V, $r_{o2} = r_{o4} = 100$ kΩ, $r_{o6} = r_{o7} = 40$ kΩ, and $C_2 = 1$ pF.

(a) Find the value of C_C that results in $f_t = 100$ MHz. What is the 3-dB frequency of the open-loop gain?

(b) Find the value of the resistance R that when placed in series with C_C causes the transmission zero to be located at infinite frequency.

(c) Find the frequency of the second pole and hence find the excess phase lag at $f = f_t$, introduced by the second pole, and the resulting phase margin assuming that the situation in (b) pertains.

Ans. 1.6 pF; 50 kHz; 500 Ω; 318 MHz; 17.4°; 72.6°

13.1.6 Slew Rate

We discussed the slew-rate limitation of op amps in Chapter 2. Here, we illustrate the origin of the slewing phenomenon in the context of the two-stage CMOS amplifier under study.

Consider the unity-gain follower of Fig. 13.6 with a step of, say, 1 V applied at the input. Because of the amplifier dynamics, its output will not change in zero time. Thus, immediately after the input is applied, the entire value of the step will appear as a differential signal between the two input terminals. In all likelihood, such a large signal will exceed the voltage required to turn off one side of the input differential pair ($\sqrt{2}V_{OV1}$: see earlier illustration in Chapter 9, Fig. 9.6) and switch the entire bias current I to the other side. Fig. 13.1 shows that for our example, Q_2 will turn off, and Q_1 will conduct the entire current I. Thus Q_4 will sink a current I that will be pulled from C_C, as shown in Fig. 13.7. Here, as we did in Fig. 13.3, we are modeling the second stage as an ideal integrator. We see that the output voltage will be a ramp with a slope of I/C_C:

$$v_O(t) = \frac{I}{C_C}t$$

Thus the slew rate, SR, is given by

$$SR = \frac{I}{C_C} \tag{13.40}$$

We should point out, however, that this is a rather simplified model of the slewing process.

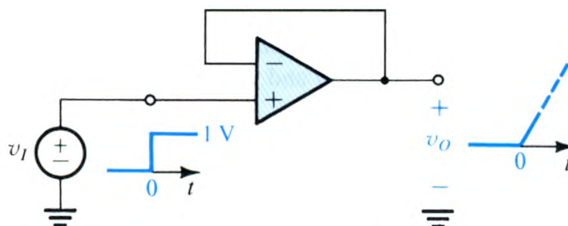


Figure 13.6 A unity-gain follower with a large step input. Since the output voltage cannot change immediately, a large differential voltage appears between the op-amp input terminals.

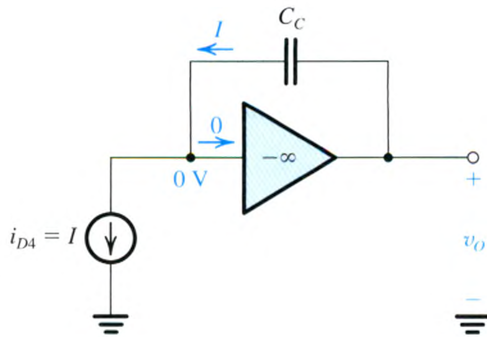


Figure 13.7 Model of the two-stage CMOS op-amp of Fig. 13.1 when a large differential voltage is applied.

For instance, we have assumed that I_{D7} is large enough to supply both the current I to C_C and the current $SR \cdot C_2$ to the output capacitance. Moreover, many op amps have mechanisms that dynamically increase their bias currents during slew rate limiting.

Relationship Between SR and f_t A simple relationship exists between the unity-gain bandwidth f_t and the slew rate SR. We find this relationship by combining Eqs. (13.31) and (13.40) and noting that $G_{m1} = g_{m1} = I/|V_{OV1}|$, to obtain

$$SR = 2\pi f_t |V_{OV1}| \tag{13.41}$$

or equivalently,

$$SR = |V_{OV1}|\omega_t \tag{13.42}$$

Thus, for a given ω_t, the slew rate is determined by the overdrive voltage at which the first-stage transistors are operated. We get a higher slew rate by operating Q_1 and Q_2 at a larger V_{OV}.

EXERCISE

13.6 Find SR for the CMOS op amp of Fig. 13.1 for the case $f_t = 100$ MHz and $V_{OV1} = 0.2$ V. If $C_C = 1.6$ pF, what must the bias current I be?
Ans. 126 V/μs; 200 μA

13.1.7 Power-Supply Rejection Ratio (PSRR)

CMOS op amps are usually used in what are known as **mixed-signal circuits**: IC chips that combine analog and digital circuits. In such circuits, the switching activity in the digital portion usually results in increased ripple on the power supplies. A portion of the supply ripple can make its way to the op-amp output and corrupt the output signal. The traditional approach for reducing supply ripple by connecting large capacitances between the supply rails and ground is not viable in IC design, as such capacitances would consume most of the

chip area. Instead, the analog IC designer has to pay attention to another op-amp specification that we have so far ignored, namely, the power-supply rejection ratio (PSRR).

The PSRR is defined as the ratio of the amplifier differential gain to the gain experienced by a change in the power-supply voltage (v_{dd} and v_{ss}). For circuits using two power supplies,

$$\mathrm{PSRR}^{+} \equiv \frac{A_d}{A^{+}} \tag{13.43}$$

and

$$\mathrm{PSRR}^{-} = \frac{A_d}{A^{-}} \tag{13.44}$$

where

$$A^{+} \equiv \frac{v_o}{v_{dd}} \tag{13.45}$$

$$A^{-} = \frac{v_o}{v_{ss}} \tag{13.46}$$

Obviously, to minimize the effect of the power-supply ripple, we need the op amp to have a large PSRR.

A detailed analysis of the PSRR of the two-stage CMOS op amp is beyond the scope of this book (see Gray et al., 2009), but we offer the following brief remarks. The circuit is remarkably insensitive to variations in V_{DD}, and thus PSRR^{+} is very high. This is not the case, however, for the negative-supply ripple v_{ss}, which is coupled to the output primarily through the second-stage transistors Q_6 and Q_7. In particular, the portion of v_{ss} that appears at the op-amp output is determined by the voltage divider formed by the output resistances of Q_6 and Q_7,

$$v_o = v_{ss}\frac{r_{o7}}{r_{o6}+r_{o7}} \tag{13.47}$$

Thus,

$$A^{-} \equiv \frac{v_o}{v_{ss}} = \frac{r_{o7}}{r_{o6}+r_{o7}} \tag{13.48}$$

Now using A_d from Eq. (13.17) gives

$$\mathrm{PSRR}^{-} \equiv \frac{A_d}{A^{-}} = g_{m1}(r_{o2} \parallel r_{o4})g_{m6}r_{o6} \tag{13.49}$$

Thus, PSRR^{-} is of the form $(g_m r_o)^2$, transistor intrinsic gain squared, and therefore is maximized by selecting long channels, L, and operating at low $|V_{OV}|$.

13.1.8 Design Trade-Offs

The performance parameters of the two-stage CMOS amplifier are primarily determined by two design parameters:

1. The length L used for the channel of each MOSFET.
2. The overdrive voltage $|V_{OV}|$ at which each transistor is operated.

So far, we have found that a larger L and correspondingly larger $|V_A|$ increases the amplifier gain, CMRR, and PSRR. We also found that operating at a lower $|V_{OV}|$ increases these three

parameters as well as increasing the input common-mode range and the allowable range of output swing. Also, although we have not analyzed the offset voltage of the op amp here, we know from Section 9.4.1 that a number of the components of the input offset voltage that arises from random device mismatches are proportional to $|V_{OV}|$ at which the MOSFETs of the input differential pair are operated. Thus the offset is minimized by operating at a lower $|V_{OV}|$. However, Eq. (13.42) illustrates that a low value of $|V_{OV1,2}|$ reduces slew rate. Additionally, too-large gate lengths will cause transistor parasitic capacitances to dominate C_1 and C_2, thus reducing unity-gain frequency.

In summary, the selection of $|V_{OV}|$ and L presents the designer with trade-offs between improving the low-frequency performance on the one hand and the high-frequency performance on the other. For modern submicron technologies, which require operation from power supplies of 1 V to 1.5 V, overdrive voltages between 0.1 V and 0.3 V are typical, and channel lengths at least 1.5 to 2 times the specified value of $L_{\min}$ are used.

Once overdrive voltages and gate lengths are chosen, all that remains is to select the transistor widths, or, equivalently, the bias currents I and $I_{D6,7}$. These bias currents provide a trade-off between unity-gain frequency and power consumption, $(V_{DD}+V_{SS})(I+I_{D6,7})$. Increasing bias current I_{D6} with fixed $V_{OV,6}$ increases $G_{m2}=g_{m6}$ [Eq. (13.12)], ω_Z [Eq. (13.26)], and ω_{P2} [Eq. (13.30)]. This, in turn, allows ω_{P1} and ω_t to be increased without compromising phase margin. We may realize these further changes either by decreasing C_C or increasing bias current I. However, we may not increase op amp bandwidth indefinitely. Increasing transistor bias current while maintaining fixed overdrive voltage and gate length entails increasing gate width, W, so that transistor parasitic capacitances will ultimately predominate over C_1 and C_2, preventing further bandwidth increases.

Example 13.1

We conclude our study of the two-stage CMOS op amp with a design example. We are required to design the circuit to obtain a dc gain of 1000 V/V. Assume that the available fabrication technology is of the 0.3-μm type for which $V_{tn}=|V_{tp}|=0.5$ V, $k_n'=200\ \mu\text{A/V}^2$, $k_p'=80\ \mu\text{A/V}^2$, $V_{An}'=|V_{Ap}'|=8$ V/μm, and $V_{DD}=V_{SS}=1.65$ V. To achieve a reasonable dc gain per stage, use $L=1\ \mu$m for all devices. Also, for simplicity, operate all devices at the same $|V_{OV}|$. Use $I=200\ \mu$A, and $I_{D6}=0.5$ mA. Specify the W/L ratios for all transistors. Also give the values realized for the input common-mode range, the maximum possible output swing, R_{in}, and R_o. Also determine the CMRR and PSRR realized. If $C_1=0.1$ pF and $C_2=0.8$ pF, find the required values of C_C and the series resistance R to place the transmission zero at $s=\infty$ and to obtain the highest possible f_t consistent with a phase margin of 85° in a unity-gain configuration. Evaluate the values obtained for f_t and SR.

Solution

Using the voltage-gain expression in Eq. (13.17),

$$A_v = g_{m1}(r_{o2}\,\|\,r_{o4})g_{m6}(r_{o6}\,\|\,r_{o7})$$

Example 13.1 *continued*

$$= \frac{2(I/2)}{V_{OV}} \times \frac{1}{2} \times \frac{V_A}{(I/2)} \times \frac{2I_{D6}}{V_{OV}} \times \frac{1}{2} \times \frac{V_A}{I_{D6}}$$

$$= \left(\frac{V_A}{V_{OV}}\right)^2$$

To obtain $A_v = 1000$, given $V_A = 8$ V,

$$1000 = \frac{64}{V_{OV}^2} \Rightarrow V_{OV} = 0.253 \text{ V}$$

To obtain the required (W/L) ratios of Q_1 and Q_2,

$$I_{D1} = \frac{1}{2}k_p'\left(\frac{W}{L}\right)_1 V_{OV}^2$$

$$100 = \frac{1}{2} \times 80\left(\frac{W}{L}\right)_1 \times 0.253^2$$

Thus,

$$\left(\frac{W}{L}\right)_1 = \left(\frac{W}{L}\right)_2 = \frac{39\ \mu\text{m}}{1\ \mu\text{m}}$$

For Q_3 and Q_4, we write

$$100 = \frac{1}{2} \times 200\left(\frac{W}{L}\right)_3 \times 0.253^2$$

to obtain

$$\left(\frac{W}{L}\right)_3 = \left(\frac{W}{L}\right)_4 = \frac{15.6\ \mu\text{m}}{1\ \mu\text{m}}$$

For Q_5,

$$200 = \frac{1}{2} \times 80\left(\frac{W}{L}\right)_5 \times 0.253^2$$

Thus,

$$\left(\frac{W}{L}\right)_5 = \frac{78\ \mu\text{m}}{1\ \mu\text{m}}$$

Since Q_7 is required to conduct 500 μA, its (W/L) ratio should be 2.5 times that of Q_5,

$$\left(\frac{W}{L}\right)_7 = 2.5\left(\frac{W}{L}\right)_5 = \frac{195\ \mu\text{m}}{1\ \mu\text{m}}$$

For Q_6, we write

$$500 = \frac{1}{2} \times 200 \times \left(\frac{W}{L}\right)_6 \times 0.253^2$$

Thus,

$$\left(\frac{W}{L}\right)_6 = \frac{78\ \mu\text{m}}{1\ \mu\text{m}}$$

At this point we should check that condition (13.1) is satisfied, which is indeed the case, ensuring that there will be no systematic output offset voltage.

Finally, let's select $I_{\text{REF}} = 20\ \mu\text{A}$, thus

$$\left(\frac{W}{L}\right)_8 = 0.1\left(\frac{W}{L}\right)_5 = \frac{7.8\ \mu\text{m}}{1\ \mu\text{m}}$$

The input common-mode range can be found using the expression in Eq. (13.4) as

$$-1.40\ \text{V} \le V_{ICM} \le 0.64\ \text{V}$$

The maximum signal swing allowable at the output is found using the expression in Eq. (13.5) as

$$-1.40\ \text{V} \le v_O \le 1.40\ \text{V}$$

The input resistance is practically infinite, and the output resistance is

$$R_o = r_{o6} \,\|\, r_{o7} = \frac{1}{2} \times \frac{8}{0.5} = 8\ \text{k}\Omega$$

The CMRR is determined using Eq. (13.19),

$$\text{CMRR} = g_{m1}(r_{o2} \,\|\, r_{o4})(2g_{m3}R_{SS})$$

where $R_{SS} = r_{o5} = V_A/I$. Thus,

$$\text{CMRR} = \frac{2(I/2)}{V_{OV}} \times \frac{1}{2} \times \frac{V_A}{(I/2)} \times 2 \times \frac{2(I/2)}{V_{OV}} \times \frac{V_A}{I}$$

$$= 2\left(\frac{V_A}{V_{OV}}\right)^2 = 2\left(\frac{8}{0.253}\right)^2 = 2000$$

Expressed in decibels, we have

$$\text{CMRR} = 20 \log 2000 = 66\ \text{dB}$$

The PSRR is determined using Eq. (13.49):

$$\text{PSRR} = g_{m1}(r_{o2} \,\|\, r_{o4})g_{m6}r_{o6}$$

Example 13.1 *continued*

$$= \frac{2(I/2)}{V_{OV}} \times \frac{1}{2} \times \frac{V_A}{(I/2)} \times \frac{2I_{D6}}{V_{OV}} \times \frac{V_A}{I_{D6}}$$

$$= 2\left(\frac{V_A}{V_{OV}}\right)^2 = 2\left(\frac{8}{0.253}\right)^2 = 2000$$

or, expressed in decibels,

$$\text{PSRR} = 20\log 2000 = 66 \text{ dB}$$

To determine f_{P2} we use Eq. (13.30) and substitute for G_{m2},

$$G_{m2} = g_{m6} = \frac{2I_{D6}}{V_{OV}} = \frac{2 \times 0.5}{0.253} = 3.95 \text{ mA/V}$$

Thus,

$$f_{P2} = \frac{3.95 \times 10^{-3}}{2\pi \times 0.8 \times 10^{-12}} = 786 \text{ MHz}$$

To move the transmission zero to $s = \infty$, we select the value of R as

$$R = \frac{1}{G_{m2}} = \frac{1}{3.95 \times 10^{-3}} = 253\ \Omega$$

For a phase margin of 85°, the phase shift due to the second pole at $f = f_t$ must be 5°, that is,

$$\tan^{-1}\frac{f_t}{f_{P2}} = 5°$$

Thus,

$$f_t = 786 \times \tan 5° = 68.7 \text{ MHz}$$

The value of C_C can be found using Eq. (13.31),

$$C_C = \frac{G_{m1}}{2\pi f_t}$$

where

$$G_{m1} = g_{m1} = \frac{2 \times 100\ \mu\text{A}}{0.253 \text{ V}} = 0.79 \text{ mA/V}$$

Thus,

$$C_C = \frac{0.79 \times 10^{-3}}{2\pi \times 68.7 \times 10^6} = 1.8 \text{ pF}$$

The value of SR can now be found using Eq. (13.41) as

$$SR = 2\pi \times 68.7 \times 10^6 \times 0.253 = 109 \text{ V/}\mu\text{s}$$

EXERCISES

13.7 A two-stage CMOS op amp is used in a unity-gain feedback configuration with a load capacitance $C_2 = 15$ pF, $I = 0.3$ mA, and $I_{D6,7} = 0.5$ mA. All overdrive voltages are $|V_{OV}| = 0.25$ V, and gate lengths are $L = 0.4$ µm. Estimate f_{P2} and then find C_C to provide 70 degrees phase margin. (You may assume a resistor is added in series with C_C to eliminate the effect of ω_z.)
Ans. 42 MHz; 12.5 pF

13.8 Assume all transistors in Fig. 13.1 have the parameters of the 0.18-µm CMOS technology in Appendix K and all gate lengths are $L = 0.6$ μm. The PMOS current-mirror overdrive voltages are $|V_{OV5,7,8}| = 0.5$ V. Find transistor widths W_5, W_7, and W_8 so that $I = I_{D6,7}/2 = 10I_{REF} = 0.2$ mA. Replace the current source I_{REF} with a resistor of the appropriate value assuming $V_{DD} = V_{SS} = 1$ V.
Ans. 11 µm; 22 µm; 1.1 µm; 50 kΩ

13.2 The Folded-Cascode CMOS Op Amp

In this section, we look at another type of CMOS op-amp circuit: the folded cascode. The circuit is a modification of the cascode amplifier studied in Section 8.5.1. A differential version of the NMOS cascode circuit in Fig. 8.31 would require stacking a large number of transistors, which presents a problem with low-voltage power supplies. Instead, we can use a PMOS transistor for the CG stage Q_2 while keeping an NMOS transistor for the CS device Q_1, as shown in Fig. 13.8(a). We need an additional current source, I_2, to bias Q_2 and serve as its active load. Note that Q_1 operates at a bias current of $(I_1 - I_2)$. Finally, we need a dc voltage V_{G2} at the gate of cascode transistor Q_2. We select its value so that Q_2 and Q_1 operate in the saturation region.

The small-signal operation of the circuit in Fig. 13.8(a) is similar to that of the NMOS cascode, and is modeled in Fig. 13.8(b). The difference here is that the signal current $g_m v_i$ is "folded down" and made to flow into the source terminal of Q_2, which is why the circuit is called a **folded cascode**.[5] As before, $R_o \approx g_{m2} r_{o2} r_{o1}$. As suggested by the model in Fig. 13.8(b), the folded cascode is a single-stage amplifier. The dominant pole will be determined by the relatively large output resistance, R_o, and the load capacitance. By contrast, in the two-stage op amp, the load capacitance determined the *second* pole frequency, and the dominant pole must be made significantly lower to ensure adequate phase margin. Hence, the folded cascode op amp can be designed to provide performance that in some respects exceeds that of the two-stage topology, particularly when driving predominantly capacitive loads such as the input of other CMOS circuits within the same integrated circuit. The folded cascode is also a popular choice for discrete op amps when followed by an output stage.[6]

[5] In this same vein, the regular cascode is sometimes referred to as a *telescopic cascode* because the stacking of transistors resembles the extension of a telescope.

[6] Examples include the ADA4661 and AD8531 op amps from Analog Devices.

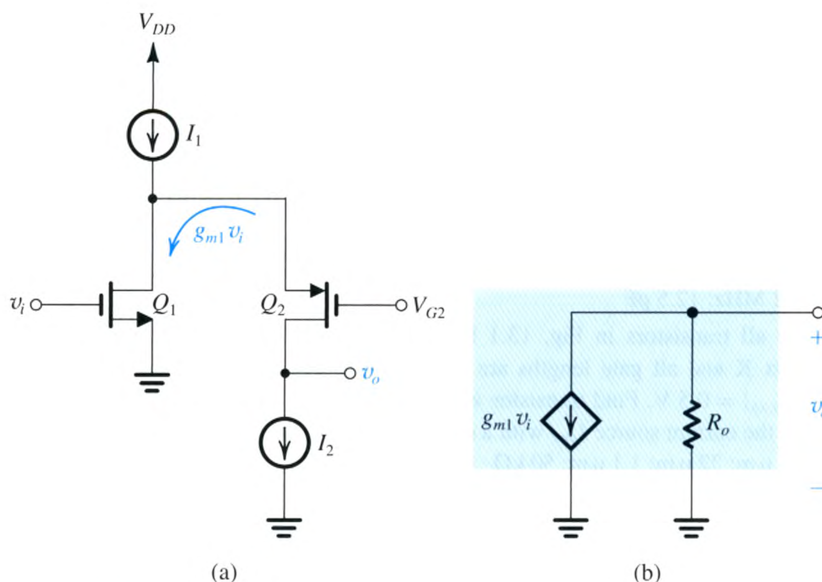


Figure 13.8 (**a**) The folded cascode amplifier with ideal current-source load. (**b**) An equivalent small-signal representation.

13.2.1 The Circuit

Figure 13.9 shows the structure of the CMOS folded-cascode op amp. Here, Q_1 and Q_2 form the input differential pair, and Q_3 and Q_4 are the cascode transistors. Recall that for differential input signals, each of Q_1 and Q_2 acts as a common-source amplifier. Also note that the gate terminals of Q_3 and Q_4 are connected to a constant dc voltage (V_{BIAS1}) and hence are at signal ground. Thus, for differential input signals, each of the transistor pairs Q_1–Q_3 and Q_2–Q_4 acts as a folded-cascode amplifier, such as the one in Fig. 13.8. Note that the input differential pair is biased by a constant-current source I. Thus both Q_1 and Q_2 operate at a bias current $I/2$. A node equation at each of their drains shows that the bias current of each of Q_3 and Q_4 is $(I_B - I/2)$.

EXERCISE

13.9 To limit the power dissipation in the op-amp circuit of Fig. 13.9 to an acceptable level, the total dc current is limited to 0.3 mA. If it is desired to bias each of Q_1 and Q_2 at a dc current four times the bias current of each of Q_3 and Q_4, find the values of I_B, I, $I_{D1,2}$, and $I_{D3,4}$.
Ans. 150 μA; 240 μA; 120 μA; 30 μA

As we learned in Chapter 8, to realize the full advantage of the high output resistance achieved through cascoding, we must make the output resistance of the current-source load equally high. This is why we use the cascode current mirror Q_5 to Q_8 in the circuit

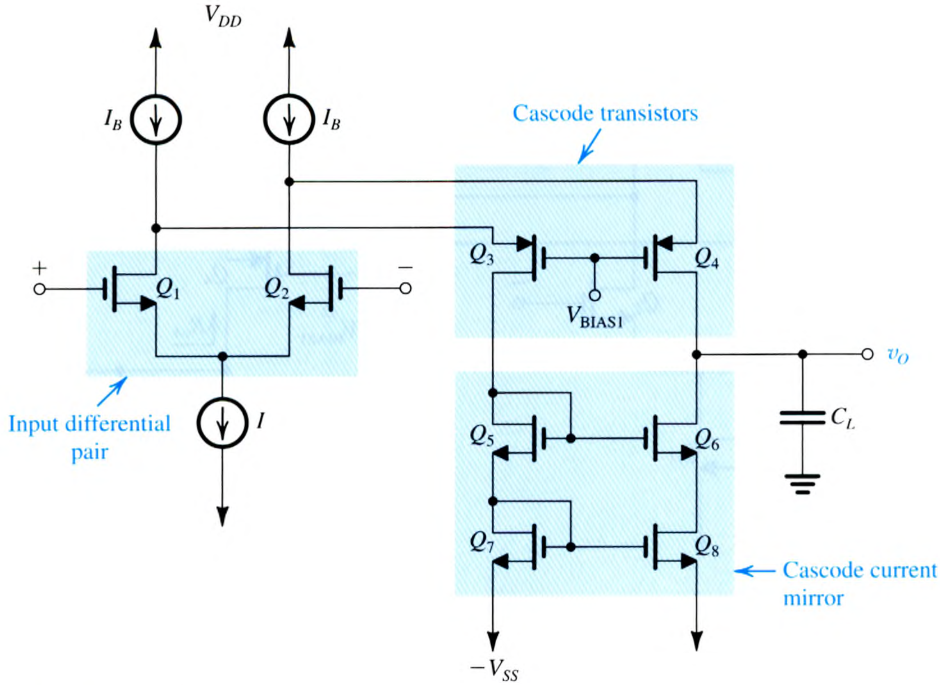


Figure 13.9 Structure of the folded-cascode CMOS op amp.

of Fig. 13.9. (We studied this current-mirror circuit in Section 8.7.1.) Finally, note that capacitance C_L denotes the total capacitance at the output node. It includes the internal transistor capacitances, an actual load capacitance (if any), and possibly an additional capacitance deliberately introduced for frequency compensation. In many cases, however, the load capacitance will be large enough that we do not need to provide additional capacitance to achieve the desired frequency compensation. We will discuss this topic shortly. For now, note that unlike the two-stage circuit, which requires a separate compensation capacitor C_C, here the load capacitance contributes to frequency compensation.

A more complete circuit for the CMOS folded-cascode op amp is shown in Fig. 13.10. Here we show the two transistors Q_9 and Q_{10}, which provide the constant bias currents I_B, and transistor Q_{11}, which provides the constant current I used for biasing the differential pair. Notice that we're not showing the details for generating the bias voltages V_{BIAS1}, V_{BIAS2}, and V_{BIAS3}. Nevertheless, we are interested in how these voltages should be selected. To that end, we next evaluate the input common-mode range and the allowable output swing.

13.2.2 Input Common-Mode Range and Output Swing

To find the input common-mode range, let the two input terminals be tied together and connected to a voltage V_{ICM}. The maximum value of V_{ICM} is limited by the requirement that Q_1 and Q_2 operate in saturation at all times. Thus $V_{ICM\max}$ should be at most V_{tn} volts above the voltage at the drains of Q_1 and Q_2. The latter voltage is determined by V_{BIAS1}, which we must choose to allow for a voltage drop across Q_9 and Q_{10} at least equal to their overdrive voltage, $|V_{OV9}| = |V_{OV10}|$. Assuming that we are able to select V_{BIAS1} so that Q_9 and Q_{10} are

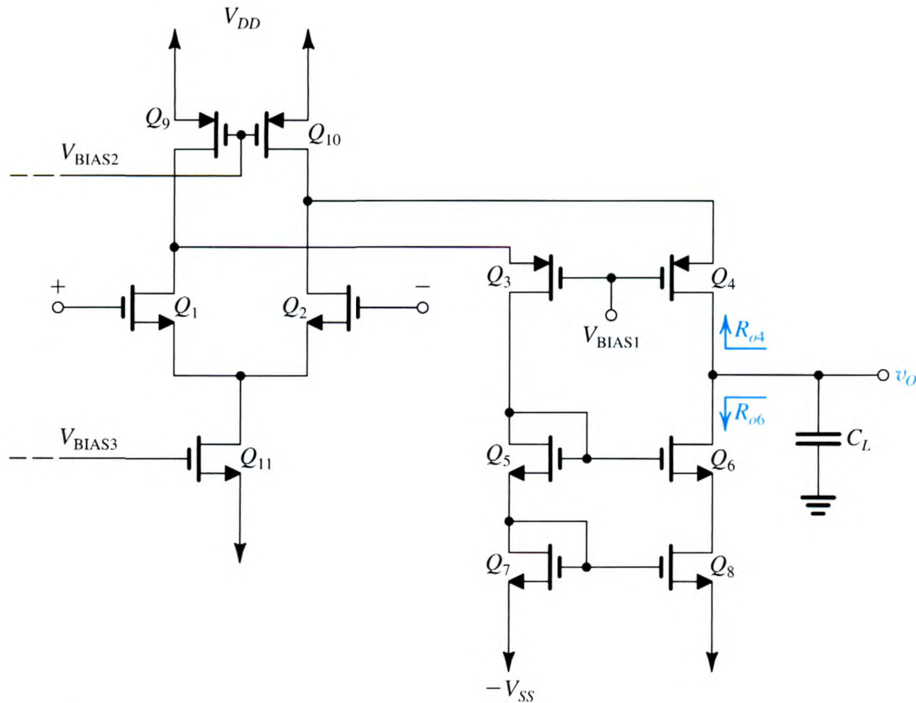


Figure 13.10 A more complete circuit for the folded-cascode CMOS amplifier of Fig. 13.9.

indeed operated at the edge of saturation, V_{ICMmax} will be

$$V_{ICM\text{max}} = V_{DD} - |V_{OV9}| + V_{tn} \tag{13.50}$$

which can be larger than V_{DD}, a significant improvement over the case of the two-stage circuit. We select the value of V_{BIAS2} to yield the required value of I_B while operating Q_9 and Q_{10} at a small value of $|V_{OV}|$ (e.g., 0.2 V or so). The minimum value of V_{ICM} is limited by the need to keep Q_{11} operating in saturation at all times, which is assured by keeping the voltage across it no smaller than V_{OV11} at all times. Thus

$$V_{ICM\text{min}} = -V_{SS} + V_{OV11} + V_{OV1} + V_{tn} \tag{13.51}$$

The presence of the threshold voltage V_{tn} in this expression indicates that $V_{ICM\text{min}}$ is not low enough for many applications. Later in this section, we describe an ingenious technique for solving this problem. For the time being, note that we should select a value for V_{BIAS3} that will provide the required value of I while operating Q_{11} at a low overdrive voltage. Combining Eqs. (13.50) and (13.51) provides

$$-V_{SS} + V_{OV11} + V_{OV1} + V_{tn} \leq V_{ICM} \leq V_{DD} - |V_{OV9}| + V_{tn} \tag{13.52}$$

The upper end of the allowable range of v_O is determined by the need to maintain Q_{10} and Q_4 in saturation. Note that Q_{10} will operate in saturation as long as an overdrive voltage, $|V_{OV10}|$, appears across it. It follows that to maximize the allowable positive swing of v_O (and also $V_{ICM\text{max}}$), we should select the value of V_{BIAS1} so that Q_{10} operates at the edge of

saturation, that is,

$$V_{\text{BIAS1}} = V_{DD} - |V_{OV10}| - V_{SG4} \tag{13.53}$$

The upper limit of v_O will then be

$$v_{O\max} = V_{DD} - |V_{OV10}| - |V_{OV4}| \tag{13.54}$$

which is two overdrive voltages below V_{DD}. The situation is not as good, however, at the other end: Since the voltage at the gate of Q_6 is $-V_{SS} + V_{GS7} + V_{GS5}$ or equivalently $-V_{SS} + V_{OV7} + V_{OV5} + 2V_{tn}$, the lowest possible v_O is obtained when Q_6 reaches the edge of saturation, namely, when v_O decreases below the voltage at the gate of Q_6 by V_{tn}, that is,

$$v_{O\min} = -V_{SS} + V_{OV7} + V_{OV5} + V_{tn} \tag{13.55}$$

Note that this value is two overdrive voltages *plus* a threshold voltage above $-V_{SS}$. This is a drawback of using the cascode mirror. As we will see in Section 13.2.7, we can solve the problem by using a modified mirror circuit.

EXERCISE

13.10 For a particular design of the folded-cascode op amp of Fig. 13.10, ±1.65-V supplies are used and all transistors are operated at overdrive voltages of 0.3-V magnitude. The fabrication process provides $V_{tn} = |V_{tp}| = 0.5$ V. Find the input common-mode range and the range allowed for v_O.
Ans. –0.55 V to +1.85 V; –0.55 V to +1.05 V

13.2.3 Voltage Gain

The folded-cascode op amp is simply a transconductance amplifier with an infinite input resistance, a transconductance G_m, and an output resistance R_o. In addition, G_m is equal to g_m of each of the two transistors of the differential pair,

$$G_m = g_{m1} = g_{m2} \tag{13.56}$$

Thus,

$$G_m = \frac{2(I/2)}{V_{OV1}} = \frac{I}{V_{OV1}} \tag{13.57}$$

The output resistance R_o is the parallel equivalent of the output resistance of the cascode amplifier and the output resistance of the cascode mirror, thus

$$R_o = R_{o4} \parallel R_{o6} \tag{13.58}$$

Figure 13.10 shows that the resistance R_{o4} is the output resistance of the CG transistor Q_4. The latter has a resistance $(r_{o2} \parallel r_{o10})$ in its source lead, thus

$$R_{o4} \simeq (g_{m4}r_{o4})(r_{o2} \parallel r_{o10}) \tag{13.59}$$

The resistance R_{o6} is the output resistance of the cascode mirror and is given by Eq. (8.97), thus

$$R_{o6} \simeq g_{m6} r_{o6} r_{o8} \tag{13.60}$$

Combining Eqs. (13.58) to (13.60) gives

$$R_o = [g_{m4} r_{o4} (r_{o2} \| r_{o10})] \| (g_{m6} r_{o6} r_{o8}) \tag{13.61}$$

We can now find the dc open-loop gain using G_m and R_o, as

$$A_v = G_m R_o \tag{13.62}$$

Thus,

$$A_v = g_{m1} \{[g_{m4} r_{o4} (r_{o2} \| r_{o10})] \| (g_{m6} r_{o6} r_{o8})\} \tag{13.63}$$

Figure 13.11 shows the equivalent-circuit model including the load capacitance C_L, which we'll take into account shortly.

Because the folded-cascode op amp is a transconductance amplifier, we call it an **operational transconductance amplifier (OTA)**. Its very high output resistance, which is of the order of $g_m r_o^2$ (see Eq. 13.61) allows it to realize a relatively high voltage gain in a single amplifier stage. You may be concerned by such a high output resistance; after all, in Chapter 2, we said that an ideal op amp has a zero output resistance! To alleviate this concern somewhat, let's find the closed-loop output resistance of a unity-gain follower formed by connecting the output terminal of the circuit of Fig. 13.10 back to the negative input terminal. Since this feedback is of the voltage sampling type, it reduces the output resistance by the factor $(1 + A\beta)$, where $A = A_v$ and $\beta = 1$, that is,

$$R_{of} = \frac{R_o}{1 + A_v} \simeq \frac{R_o}{A_v} \tag{13.64}$$

Substituting for A_v from Eq. (13.62) gives

$$R_{of} \simeq \frac{1}{G_m} \tag{13.65}$$

which is a general result that applies to any OTA to which we apply 100% voltage feedback. For our particular circuit, $G_m = g_{m1}$, thus

$$R_{of} = 1/g_{m1} \tag{13.66}$$

Since g_{m1} is of the order of 1 mA/V, R_{of} will be of the order of 1 kΩ. Although this is not very small, it is reasonable in view of the simplicity of the op-amp circuit as well as the fact that this type of op amp is not usually intended to drive low-valued resistive loads.

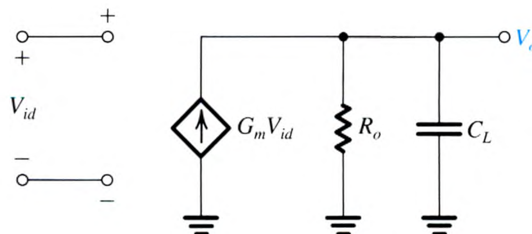


Figure 13.11 Small-signal equivalent circuit of the folded-cascode CMOS amplifier. Note that this circuit is in effect an operational transconductance amplifier (OTA).

EXERCISE

13.11 The CMOS op amp of Figs. 13.9 and 13.10 is fabricated in a process for which $V'_{An} = |V'_{Ap}| = 20$ V/μm. If all devices have 1-μm channel length and are operated at equal overdrive voltages of 0.2-V magnitude, $I = 240$ μA, and $I_B = 150$ μA, find the voltage gain and the value of R_o obtained.
Ans. 16,000 V/V; 13.3 MΩ

13.2.4 Frequency Response

From Section 10.5, we know that one of the advantages of the cascode configuration is its excellent high-frequency response. It has three poles: one at the input, one at the connection between the CS and CG transistors (i.e., at the source terminals of Q_3 and Q_4), and one at the output terminal. Normally, the first two poles are at very high frequencies, especially when the resistance of the signal generator that feeds the differential pair is small. Since the primary purpose of CMOS op amps is to feed capacitive loads, C_L is usually large, and the pole at the output becomes dominant. Even if C_L is not large, we can increase it deliberately to give the op amp a dominant pole and thus an open-loop gain that decreases at the uniform rate of −20 dB/decade down to the unity-gain frequency f_t, ensuring stable operation when feedback is applied. From Fig. 13.11 we can write

$$\frac{V_o}{V_{id}} = \frac{G_m R_o}{1 + sC_L R_o} \tag{13.67}$$

Thus, the dominant pole has a frequency f_P,

$$f_P = \frac{1}{2\pi C_L R_o} \tag{13.68}$$

and the unity-gain frequency f_t will be

$$f_t = G_m R_o f_P = \frac{G_m}{2\pi C_L} \tag{13.69}$$

From a design point of view, the value of C_L should be such that at $f = f_t$ the excess phase resulting from the nondominant poles is small enough that we can achieve the required phase margin. If C_L is not large enough, it can be augmented. Alternatively, we can reduce G_m by reducing the bias current in $Q_{1,2}$ while keeping $V_{OV1,2}$ constant, which has the benefit of reducing power consumption.

It is important to note how increasing the load capacitance affects the operation of each of the two op-amp circuits we have studied differently. In the two-stage circuit, if C_L is increased, the frequency of the second pole decreases, the excess phase shift at $f = f_t$ increases, and the phase margin is reduced. Here, on the other hand, when C_L is increased, f_t decreases, but the phase margin increases. In other words, a heavier capacitive load decreases the bandwidth of the folded-cascode amplifier but does not impair its response (which happens when the phase margin decreases). We can restore the lost bandwidth by increasing G_m in Eq. (13.69). We do so by increasing the bias currents in $Q_{1,2}$ while keeping $V_{OV1,2}$ constant, resulting in increased power consumption.

13.2.5 Slew Rate

As we discussed in Section 13.1.6, slewing occurs when a large differential signal appears at the op-amp input. Refer to Fig. 13.9 and consider the case when V_{id} is large and turns Q_2 off. Transistor Q_1 will then attempt to conduct the entire bias current I. This, however, would not be possible, since I is usually larger than I_B. Considering the drain node of Q_1 we see that for the node equation to be satisfied, not only must the current in Q_3 reduce to zero, but also the current of Q_1 must reduce to equal I_B. For this to happen, both Q_1 and the transistor supplying I must enter the triode mode of operation, and the voltages at their drains must fall accordingly. Now, the zero current in Q_3 causes the input current of the mirror to be zero, and correspondingly its output current, in the drain of Q_6, will be zero. Meanwhile, the zero current in the drain of Q_2 forces the entire current I_B to flow through Q_4 and into C_L. This causes the output voltage v_O to ramp with a slope of I_B/C_L, which is the slew rate,

$$SR = \frac{I_B}{C_L} \tag{13.70}$$

After the slewing process is completed, before the amplifier can return to its normal linear operation, both Q_1 and the transistor that supplies the bias current I must leave the triode mode and return to the saturation mode of operation. This, however, can take some time and may introduce additional distortion in the output signal. This is where creative circuit design again comes to the rescue! Problem 13.35 investigates an ingenious way to deal with this issue.

Example 13.2

Consider the design of the CMOS folded-cascode op amp in Fig. 13.10 with $I = 200$ μA, $I_B = 150$ μA, and $|V_{OV}| = 0.2$ V for all transistors. The transistor parameters are given in Appendix K for the 0.18-μm technology. For NMOS: $k_n' = 387$ μA/V^2 and $|V_A'| = 5$ V/μm. For PMOS: $k_p' = 86$ μA/V^2 and $|V_A'| =$ 6 V/μm. Also, $V_{DD} = 1.8$ V, $V_{SS} = 0$ V and $|V_t| = 0.5$ V. Let all transistors have $L = 0.4$ μm and assume $C_L = 2$ pF. Find I_D, g_m, r_o, and W/L for all transistors. Find the allowable range of V_{ICM} and the output voltage swing. Determine the values of A_v, f_t, f_P, and SR. What is the power dissipation of the op amp?

Solution

From the given values of I and I_B we can determine the drain current I_D for each transistor. The transconductance of each device is found using

$$g_m = \frac{2I_D}{V_{OV}} = \frac{2I_D}{0.2}$$

and the output resistance r_o from

$$r_o = \frac{|V_A|}{I_D} = \frac{|V_A'|L}{I_D}$$

The W/L ratio for each transistor is determined from

$$\left(\frac{W}{L}\right) = \frac{2I_D}{k'V_{OV}^2}$$

Transistors of the same type having the same bias current and overdrive voltage will have the same g_m, r_o, and W/L. Hence, we summarize the results as follows:

	$Q_{1,2}$	$Q_{3,4}$	$Q_{5,6,7,8}$	$Q_{9,10}$	Q_{11}
I_D (μA)	100	50	50	150	200
g_m (mA/V)	1.0	0.5	0.5	1.5	2.0
r_o (kΩ)	20	48	40	16	10
W/L	12.9	29.1	6.5	87.3	25.8

Using Eq. (13.52), we find the input common-mode range,

$$0.9\text{ V} \leq V_{ICM} \leq 2.1\text{ V}$$

We find the output voltage swing from Eqs. (13.54) and (13.55),

$$0.9\text{ V} \leq v_O \leq 1.4\text{ V}$$

To obtain the voltage gain, we first determine R_{o4} and R_{o6} using Eqs. (13.59) and (13.60),

$$R_{o4} \approx 48 \times 0.5(20 \| 16) = 213.3\text{ k}\Omega$$

$$R_{o6} \approx 0.5 \times 40 \times 40 = 800\text{ k}\Omega$$

We then find the output resistance,

$$R_o = R_{o4} \| R_{o6} = 168.4\text{ k}\Omega$$

and the voltage gain,

$$A_v = G_m R_o = 1 \times 10^{-3} \times 168.4 \times 10^3 = 168.4\text{ V/V} = 44.5\text{ dB}$$

Note that this would be a small value for a general-purpose discrete op amp. However, it is often sufficient for driving capacitive loads on the same integrated circuit as the op amp.

We find the unity-gain bandwidth using Eq. (13.69),

$$f_t = \frac{1\text{ mA/V}}{2\pi \times 2\text{ pF}} = 79.6\text{ MHz}$$

Thus, the dominant-pole frequency must be

$$f_P = \frac{f_t}{A_v} = 473\text{ kHz}$$

We can find the slew rate using Eq. (13.70),

$$SR = \frac{I_B}{C_L} = \frac{0.15\text{ mA}}{2\text{ pF}} = 75\text{ V/}\mu\text{s}$$

Finally, to determine the power dissipation, we note that the total current is 300 μA and the supply voltage is 1.8 V, so

$$P_D = 1.8 \times 0.3 = 0.54\text{ mW}$$

13.2.6 Increasing the Input Common-Mode Range: Rail-to-Rail Input Operation

In Section 13.2.2, we found that while the upper limit on the input common-mode range exceeds the supply voltage V_{DD}, the lower limit is significantly higher than $-V_{SS}$. The opposite situation occurs if the input differential amplifier is made up of PMOS transistors. It follows that an NMOS and a PMOS differential pair placed in parallel would provide an input stage with a common-mode range that exceeds the power-supply voltage in both directions. This is known as rail-to-rail input operation. Figure 13.12 shows such an arrangement. To keep the diagram simple, we have not shown the parallel connection of the two differential pairs: The two positive-input terminals are to be connected together and the two negative-input terminals are to be tied together. Transistors Q_5 and Q_6 are the cascode transistors for the Q_1–Q_2 pair, and transistors Q_7 and Q_8 are the cascode devices for the Q_3–Q_4 pair. The output voltage V_o is shown taken differentially between the drains of the cascode devices. To obtain a single-ended output, a differential-to-single-ended conversion circuit should be connected in cascade.

Figure 13.12 indicates by arrows the direction of the current increments that result from the application of a positive differential input signal V_{id}. The currents generated by the NMOS pair are $G_{mn}(V_{id}/2)$, where $G_{mn} = g_{m1} = g_{m2}$, whereas those generated in the PMOS pair

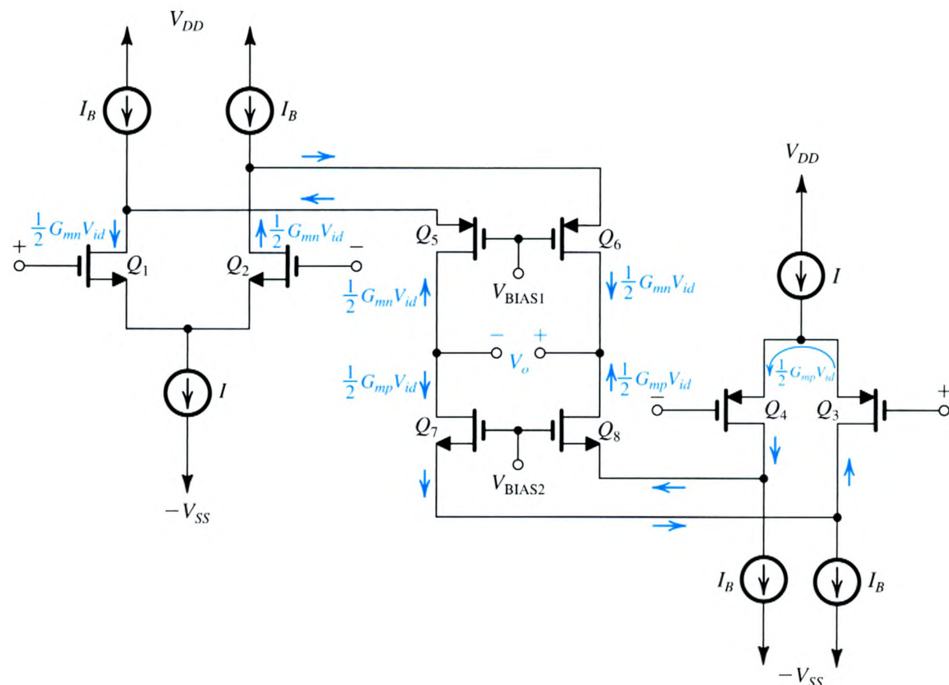


Figure 13.12 A folded-cascode op amp that employs two parallel complementary input stages to achieve rail-to-rail input common-mode operation. Note that the two "+" terminals are connected together and the two "–" terminals are connected together.

are $G_{mp}(V_{id}/2)$, where $G_{mp} = g_{m3} = g_{m4}$. Thus, the total current feeding each output node is $(G_{mn} + G_{mp})(V_{id}/2)$. If the small-signal output resistance between each of the two output nodes and ground is denoted R_o, the output voltage will be

$$V_o = (G_{mn} + G_{mp})R_o V_{id}$$

Thus, the voltage gain will be

$$A_v = (G_{mn} + G_{mp})R_o \tag{13.71}$$

This, however, assumes that both differential pairs will be operating simultaneously. This occurs only over a limited range of V_{ICM}. Over the remainder of the input common-mode range, only one of the two differential pairs will be operational, and the gain drops to either $G_{mn}R_o$ or $G_{mp}R_o$ depending on which differential pair is operational. In situations where the input common-mode can change during operation, such as in a unity-gain configuration with wide input swing, differences between the offset voltages of the two differential pairs can cause **crossover distortion**. Nevertheless, this rail-to-rail, folded-cascode structure is used in several commercially available op amps including the Texas Instruments OPA357, and Analog Devices AD8531 and ADA4661.

EXERCISE

13.12 For the circuit in Fig. 13.12, assume that all transistors, including those that implement the current sources, are operating at equal overdrive voltages of 0.3-V magnitude and have $|V_t| = 0.7$ V and that $V_{DD} = V_{SS} = 2.5$ V.

(a) Find the range over which the NMOS input stage operates.

(b) Find the range over which the PMOS input stage operates.

(c) Find the range over which both operate (the overlap range).

(d) Find the input common-mode range.

(Note that to operate properly, each of the current sources requires a minimum voltage of $|V_{OV}|$ across its terminals.)

Ans. –1.2 V to +2.9 V; –2.9 V to +1.2 V, –1.2 V to +1.2 V; –2.9 V to +2.9 V

13.2.7 Increasing the Output Voltage Range: The Wide-Swing Current Mirror

In Section 13.2.2, we found that while the output voltage of the circuit of Fig. 13.10 can swing to within $2|V_{OV}|$ of V_{DD}, the cascode current mirror limits the negative swing to $[2|V_{OV}| + V_t]$ above $-V_{SS}$. In other words, the cascode mirror reduces the voltage swing by V_t volts. This is further illustrated in Fig. 13.13(a), which shows a cascode mirror (with $V_{SS} = 0$, for simplicity) and indicates the voltages that result at the various nodes. Notice that because the voltage at the gate of Q_3 is $2V_t + 2V_{OV}$, the minimum voltage permitted at the output (while Q_3 remains saturated) is $V_t + 2V_{OV}$, hence the extra V_t. Also, notice that Q_1 is operating with a drain-to-source voltage $V_t + V_{OV}$, which is V_t volts greater than it needs to operate in saturation.

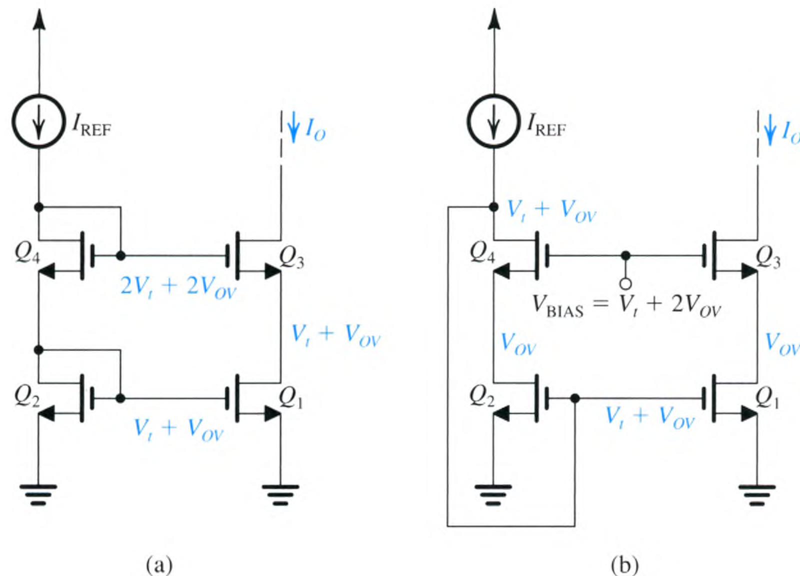


Figure 13.13 **(a)** Cascode current mirror with the voltages at all nodes indicated. Note that the minimum voltage allowed at the output is $V_t + 2V_{OV}$. **(b)** A modification of the cascode mirror that reduces the minimum output voltage to $2V_{OV}$. This is the wide-swing current mirror. The circuit requires a bias voltage V_{BIAS}.

The observations above lead us to conclude that to permit the output voltage at the drain of Q_3 to swing as low as $2V_{OV}$, we must lower the voltage at the gate of Q_3 from $2V_t + 2V_{OV}$ to $V_t + 2V_{OV}$. This is exactly what is done in the modified mirror circuit in Fig. 13.13(b): The gate of Q_3 is now connected to a bias voltage $V_{BIAS} = V_t + 2V_{OV}$. Thus the output voltage can go down to $2V_{OV}$ with Q_3 still in saturation. Also, the voltage at the drain of Q_1 is now V_{OV} and thus Q_1 is operating at the edge of saturation. The same is true of Q_2 and thus the current tracking between Q_1 and Q_2 will be assured. Note, however, that we can no longer connect the gate of Q_2 to its drain. Rather, it is connected to the drain of Q_4. This establishes a voltage of $V_t + V_{OV}$ at the drain of Q_4, which is sufficient to operate Q_4 in saturation (as long as V_t is greater than V_{OV}, which is usually the case). This circuit is known as the **wide-swing current mirror**. Finally, note that Fig. 13.13(b) does not show the circuit for generating V_{BIAS}. There are a number of possible circuits to accomplish this task, one of which is explored in Exercise 13.13.

EXERCISE

13.13 Show that if transistor Q_5 in the circuit of Fig. E13.13 has a *W/L* ratio equal to one-quarter that of the transistors in the wide-swing current mirror of Fig. 13.13(b), and if the same value of I_{REF} is used in both circuits, then the voltage generated, V_5, is $V_t + 2V_{OV}$, which is the value of V_{BIAS} needed for the gates of Q_3 and Q_4.

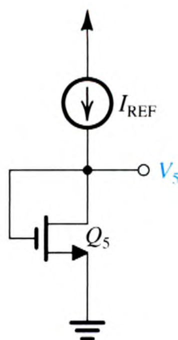


Figure E13.13

13.3 BJT Op-Amp Techniques

Although CMOS fabrication technologies have displaced bipolar technologies in many applications, op amps made with bipolar transistors have several advantages over CMOS op amps that keep them in widespread use. Bipolar transistors exhibit greater current-carrying capacity and higher breakdown voltage than MOSFETs of similar size. Hence, op amps delivering relatively large load currents or operating with large voltages are often less expensive when implemented with bipolar transistors. Bipolar transistors exhibit better matching than CMOS transistors and higher transconductance, resulting in lower amplifier offset, and they introduce less noise. Although we can overcome these shortcomings of CMOS op amps with specialized design techniques, doing so tends to increase cost. Thus, BJT op amps are particularly popular for general-purpose applications. On the other hand, the finite base current of BJTs translates into finite input bias and input offset currents and finite input resistance, nonidealities not present in CMOS op amps. Moreover, CMOS op amps can be integrated on the same chip as CMOS digital logic, which makes them preferable when an entire system combining analog and digital circuits is to be achieved on a single chip.

We begin this section by describing bias circuits for the BJT op amp. We then present common circuits for BJT op-amp input and output stages.

Device Parameters The technology we will use in the examples, exercises, and problems for this section (except for the 741 circuit in Section 13.3.4) has the following characteristics:

$$npn\text{ transistors}: \quad \beta = 40 \qquad V_A = 30\ V$$
$$pnp\text{ transistors}: \quad \beta = 10 \qquad |V_A| = 20\ V$$

For both, $|V_{BE}| \simeq 0.7$ V and $|V_{CEsat}| \simeq 0.1$ V. It is important to note that we will assume that for this technology, the transistor will remain in the active mode for $|V_{CE}|$ as low as 0.1 V (in other words, that 0.6 V is needed to forward-bias the CBJ).

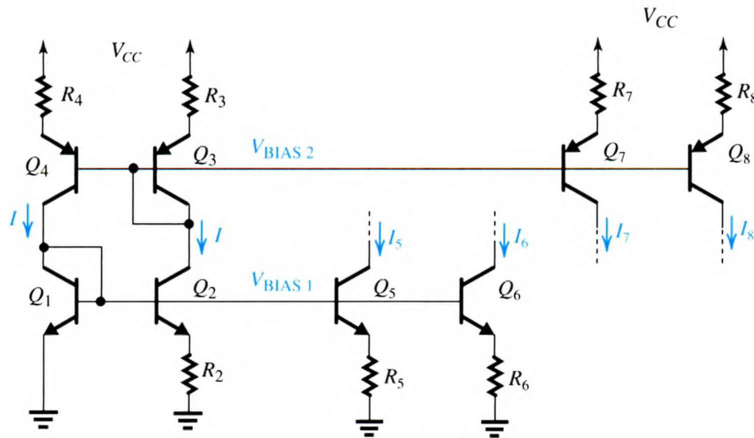


Figure 13.14 A self-biased current-reference source using a Widlar circuit to generate $I = (V_T/R_2)\ln(I_{S2}/I_{S1})$. The bias voltages V_{BIAS1} and V_{BIAS2} are used to generate bias currents I_{5-8} used elsewhere in the op amp.

13.3.1 Bias Design

As in CMOS op amps, the bias design of modern BJT amplifiers makes extensive use of current mirrors and current-steering circuits (Sections 8.2 and 8.7). Typically, however, the bias currents are small (in the microamp range). Thus, the Widlar current source (Section 8.7.4) is especially popular here. As well, emitter-degeneration resistors (in the tens-of-kilohm range) are frequently used.

Figure 13.14 shows a self-biased current-reference source that uses a Widlar circuit formed by Q_1, Q_2, and R_2, and a current mirror Q_3–Q_4 with matched emitter-degeneration resistors R_3 and R_4. The circuit establishes a current I in each of the four transistors, with the value of I determined as follows. Neglecting base currents and r_o's for simplicity, we write

$$V_{BE1} = V_T \ln\left(\frac{I}{I_{S1}}\right)$$

$$V_{BE2} = V_T \ln\left(\frac{I}{I_{S2}}\right)$$

Thus,

$$V_{BE1} - V_{BE2} = V_T \ln\left(\frac{I_{S2}}{I_{S1}}\right)$$

But,

$$V_{BE1} - V_{BE2} = IR_2$$

So,

$$I = \frac{V_T}{R_2}\ln\left(\frac{I_{S2}}{I_{S1}}\right) \tag{13.72}$$

Hence the value of I is determined by R_2 and the ratio of the emitter areas of Q_1 and Q_2. Also, notice that I is independent of V_{CC}, a highly desirable outcome. Neglecting the temperature dependence of R_2, we see that I is directly proportional to the absolute temperature. It follows

that transistors biased by I or mirrored versions of it will exhibit g_m's that are constant independent of temperature! We can use a similar circuit to maintain constant g_m's in CMOS op amps by replacing the bipolar current mirrors with CMOS ones.

EXERCISE

D13.14 Design the circuit in Fig. 13.14 to generate a current $I = 10\,\mu\text{A}$. Use transistors Q_1 and Q_2 with their areas in a 1:2 ratio. Assume that Q_3 and Q_4 are matched and design for a 0.2-V drop across each of R_3 and R_4. Specify the values of R_2, R_3, and R_4.
Ans. 1.73 kΩ; 20 kΩ; 20 kΩ

The circuit in Fig. 13.14 provides a bias line V_{BIAS1} with a voltage equal to V_{BE1}. We use this to bias Q_5 and Q_6 and generate currents proportional to I by appropriately scaling their emitter areas. Similarly, the circuit provides a bias line V_{BIAS2} at a voltage $(IR_3 + V_{EB3})$ below V_{CC}. We use this to bias Q_7 and Q_8 and generate constant currents proportional to I by appropriately scaling emitter areas and emitter-degeneration resistances.

EXERCISE

D13.15 Design the circuit in Fig. 13.14 to generate currents $I_5 = 10\,\mu\text{A}$, $I_6 = 5\,\mu\text{A}$, $I_7 = 20\,\mu\text{A}$, and $I_8 = 10\,\mu\text{A}$. Specify the required emitter areas for $Q_5 - Q_8$ as ratios of the emitter area of Q_2 and Q_3. Also specify the values required for $R_5 - R_8$. Use the values of R_2, R_3, and R_4 found in Exercise 13.14. Ignore base currents.
Ans. $A_{E5} = A_{E2}, A_{E6} = 0.5A_{E2}, A_{E7} = 2A_{E3}, A_{E8} = A_{E3}$; 1.73 kΩ, 3.46 kΩ, 10 kΩ, 20 kΩ

13.3.2 Design of the Input Stage

A two-stage op amp using bipolar transistors is shown in Fig. 13.15. The cascade of a differential pair with current-mirror load and a common-emitter stage affords high gain. The capacitor C_C is selected to provide adequate phase margin, as in the two-stage CMOS op amp. The class AB output stage, described in Chapter 12, provides a low output resistance.

Unfortunately, this circuit is unsuitable for low supply voltages. The output may swing only to within approximately 1 V of the supply voltages due to the V_{BE} drops of the output stage transistors. The alternative output stage in Section 13.3.5 affords near rail-to-rail output swing. In this section, we focus on the input stage and techniques to extend its input common-mode range. We desire rail-to-rail input common-mode range when, for example, we use the op amp in a non-inverting configuration with the input biased near one of the supply voltages. This is often necessary in circuits operating from low supply voltages.

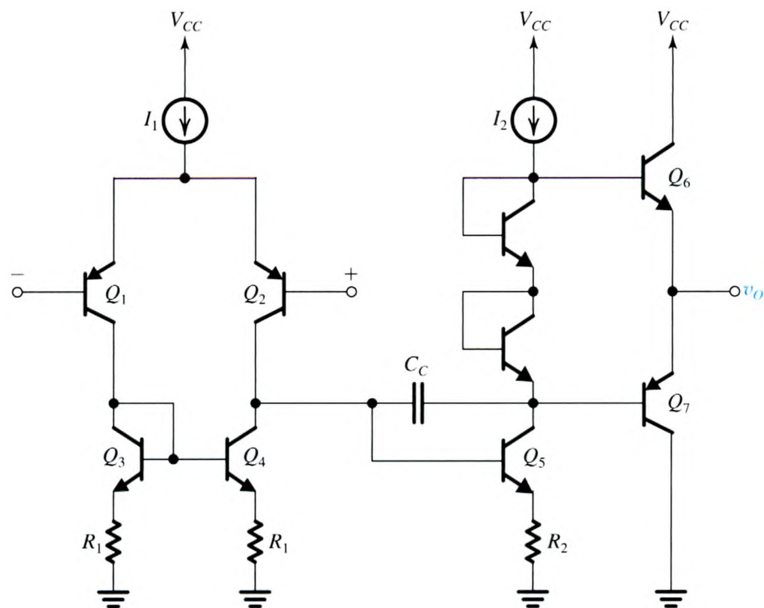


Figure 13.15 A bipolar two-stage op amp with class AB output stage. The input common-mode range is limited by the input differential pair with current-mirror load, and the output swing is limited by the class AB output stage.

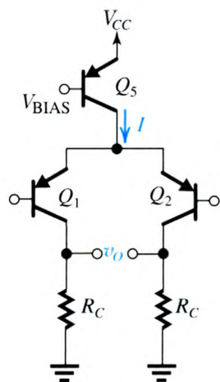


Figure 13.16 For the input common-mode range to include ground voltage, the classical current-mirror-loaded input stage in Fig. 13.15 has to be replaced with the resistively loaded configuration with the dc-voltage drop across R_C limited to 0.2 V to 0.3 V.

Consider first the low end of the input common-mode range. The value of $V_{ICM\min}$ is limited by the need to keep Q_1 in the active mode. Specifically, since the collector of Q_1 is at a voltage $V_{BE3} \simeq 0.7$ V, we see that the voltage applied to the base of Q_1 cannot go lower than 0.1 V without causing the collector–base junction of Q_1 to become forward biased. Thus $V_{ICM\min} = 0.1$ V, and the input common-mode range does *not* include ground voltage as desired.

The only way to extend $V_{ICM\min}$ to 0 V is to lower the voltage at the collector of Q_1. We can do this only by abandoning the use of the current-mirror load and using instead resistive loads, as shown in Fig. 13.16. Observe that in effect we are *going back* to the resistively loaded differential pair with which we began our study of differential amplifiers in Chapter 9!

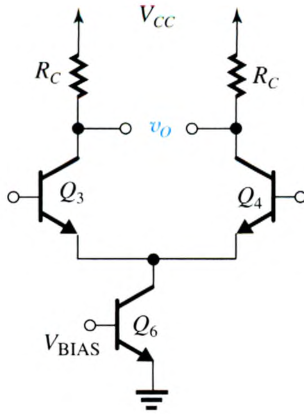


Figure 13.17 The complement of the circuit in Fig. 13.16. While the input common-mode range of the circuit in Figure 13.16 extends below ground, here it extends above V_{CC}. Connecting the two circuits in parallel, as we will show, results in a rail-to-rail V_{ICM} range.

The minimum allowed value of V_{ICM} in the circuit of Fig. 13.16 is still, of course, limited by the need to keep Q_1 and Q_2 in the active mode. This in turn is achieved by avoiding V_{ICM} values that cause the base voltages of Q_1 and Q_2 to go below their collector voltages by more than 0.6 V,

$$V_{ICM\text{min}} = V_{R_C} - 0.6 \text{ V}$$

where V_{R_C} is the voltage drop across each of R_{C1} and R_{C2}. Now if V_{R_C} is selected to be 0.2 V to 0.3 V, then $V_{ICM\text{min}}$ will be -0.4 V to -0.3 V, which is exactly what we need.

The major drawback of replacing the current-mirror load with resistive loads is that the differential gain is considerably reduced,

$$\frac{v_o}{v_{id}} = -g_{m1,2}R_C$$

$$= -\frac{I/2}{V_T}R_C = -\frac{V_{R_C}}{V_T}$$

where we have neglected r_o for simplicity. Thus for $V_{R_C} = 0.3$ V, the gain is only 12 V/V. As we will see shortly, this low-gain problem can be solved by cascoding.

Next consider the upper end of the input common-mode range. In Fig. 13.16, the maximum voltage that can be applied to the bases of Q_1 and Q_2 is limited by the need to keep the current-source transistor in the active mode. We achieve this by ensuring that the voltage across Q_5, V_{EC5}, does not fall below 0.1 V or so. Thus the maximum value of V_{ICM} will be a voltage $V_{EB1,2}$ or approximately 0.7 V lower,

$$V_{ICM\text{max}} = V_{CC} - 0.1 - 0.7 = V_{CC} - 0.8$$

That is, the upper end of the input common-mode range is at least 0.8 V below V_{CC}, a severe limitation.

To recap, while the circuit in Fig. 13.16 has $V_{ICM\text{min}}$ of a few tenths of a volt below the negative power-supply rail (at ground voltage), the upper end of V_{ICM} is rather far from V_{CC},

$$-0.3 \leq V_{ICM} \leq V_{CC} - 0.8$$

where we have assumed $V_{R_C} = 0.3$ V. To extend the upper end of V_{ICM}, we adopt a solution similar to the one we used in the CMOS case (Section 13.2.6, Fig. 13.12), namely, we use a

parallel complementary input stage. To that end, note that the *npn* version of the circuit of Fig. 13.16, shown in Fig. 13.17, has a common-input range of

$$0.8 \leq V_{ICM} \leq V_{CC} + 0.3$$

where we have assumed that $V_{R_C} = 0.3$ V. Thus, as expected, the high end meets our specifications and in fact is above the positive supply rail by 0.3 V. The lower end, however, does not, but this should cause us no concern because the lower end will be looked after by the *pnp* pair. Finally, note that there is a range of V_{ICM} in which both the *pnp* and the *npn* circuits will be active and properly operating,

$$0.8 \leq V_{ICM} \leq V_{CC} - 0.8$$

Figure 13.18 shows an input stage that achieves more than rail-to-rail input common-mode range by using a *pnp* differential pair (Q_1, Q_2) and an *npn* differential pair (Q_3, Q_4), connected in parallel. To keep the diagram simple, we are not showing the parallel connection of the input terminals; the + input terminals are assumed to be connected together, as are the – input terminals. In order to increase the gain obtained from the resistively loaded differential pairs, we add a folded-cascode stage. Here R_7 and R_8 are the resistive loads of the *pnp* pair Q_1-Q_2, and Q_7-Q_8 are its cascode transistors. Similarly, R_9 and R_{10} are the resistive loads of the *npn* pair Q_3-Q_4, and Q_9-Q_{10} are its cascode transistors. Notice that the cascode transistors do "double duty." For instance, Q_7-Q_8 operate as the cascode devices for Q_1-Q_2 and at the same time as current-source loads for Q_9-Q_{10}. We can make a similar statement about Q_9-Q_{10}. The output voltage of the first stage, v_{Od}, appears between the collectors of the cascode devices.

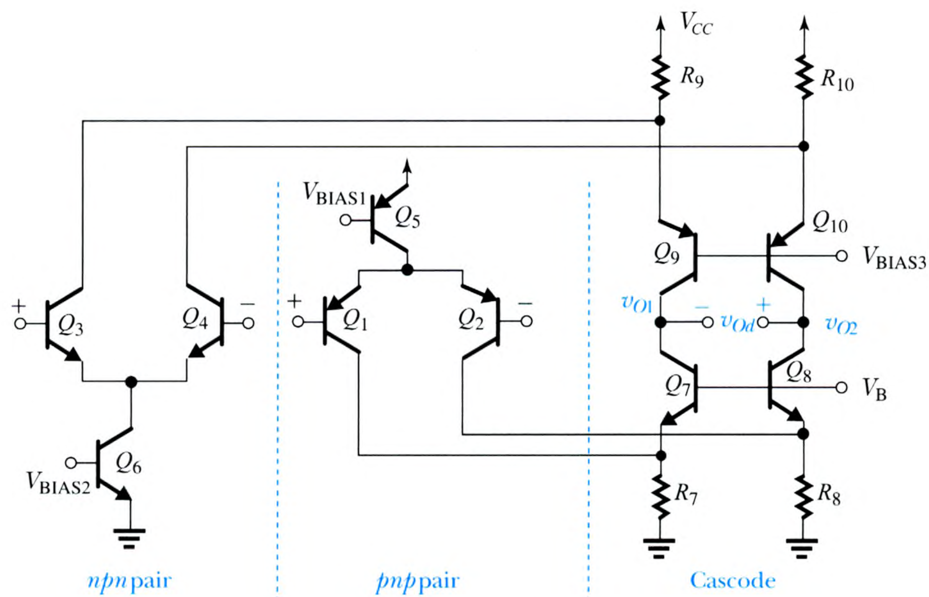


Figure 13.18 Input stage with rail-to-rail input common-mode range and a folded-cascode stage to increase the gain. Note that all the bias voltages including V_{BIAS3} and V_B are generated elsewhere on the chip.

For $V_{ICM} \ll 0.8$ V, the *npn* stage will be inactive and the gain is determined by the transconductance G_m of the Q_1–Q_2 pair together with the output resistance seen between the collectors of the cascode transistors. At the other end of V_{ICM} (that is, $V_{ICM} \gg V_{CC} - 0.8$) the Q_1–Q_2 stage will be inactive, and the gain will be determined by the transconductance G_m of the Q_3–Q_4 pair and the output resistance between the collectors of the cascode devices. In the overlap region $0.8 \leq V_{ICM} \leq V_{CC} - 0.8$, both the *pnp* and *npn* stages will be active and their effective transconductances G_m add up, resulting in a higher gain. The dependence of the differential gain on the input common-mode V_{ICM} is usually undesirable and can be reduced considerably by arranging that one of the two differential pairs is turned off when the other one is active.[7]

Example 13.3

We need to find the input resistance and the voltage gain of the input stage shown in Fig. 13.18. Let $V_{ICM} \ll 0.8$ V so that the Q_3–Q_4 pair is off. Assume that Q_5 supplies 10 μA, that each of Q_7 to Q_{10} is biased at 10 μA, and that all four cascode transistors are operating in the active mode. The input resistance of the second stage of the op amp (not shown) is $R_L = 2\,\text{M}\Omega$. The emitter-degeneration resistances are $R_7 = R_8 = 20\,\text{k}\Omega$, and $R_9 = R_{10} = 30\,\text{k}\Omega$. Recall that the device parameters are $\beta_N = 40$, $\beta_P = 10$, $V_{An} = 30$ V, $|V_{Ap}| = 20$ V.

Solution

Since the stage is fully balanced, we can use the differential half-circuit shown in Fig. 13.19(a). The input resistance R_{id} is twice the value of $r_{\pi 1}$,

$$R_{id} = 2r_{\pi 1} = 2\beta_P/g_{m1}$$

where

$$g_{m1} = \frac{I_{C1}}{V_T} = \frac{5 \times 10^{-6}}{25 \times 10^{-3}} = 0.2\ \text{mA/V}$$

Thus,

$$R_{id} = \frac{2 \times 10}{0.2} = 100\ \text{k}\Omega$$

To find the short-circuit transconductance, we short the output to ground as shown in Fig. 13.19(b) and find G_{m1} as

$$G_{m1} = \frac{i_o}{v_{id}/2}$$

[7]This is done in the NE5234 op amp, whose circuit is described and analyzed in great detail in Gray et al. (2009).

Example 13.3 *continued*

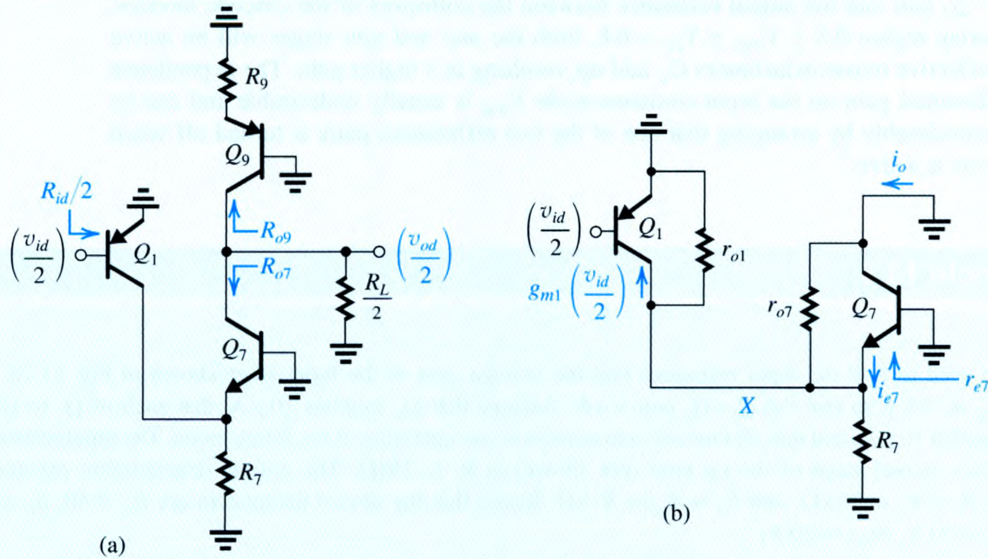


Figure 13.19 **(a)** Differential half-circuit for the input stage shown in Fig. 13.18 with $V_{ICM} \ll 0.8$ V. **(b)** Determining $G_{m1} = i_o/(v_{id}/2)$.

At node X we have four parallel resistances to ground,

$$r_{o1} = \frac{|V_{Ap}|}{I_{C1}} = \frac{20\text{ V}}{5\ \mu\text{A}} = 4\text{ M}\Omega$$

$$R_7 = 20\text{ k}\Omega$$

$$r_{o7} = \frac{V_{An}}{I_{C7}} = \frac{30\text{ V}}{10\ \mu\text{A}} = 3\text{ M}\Omega$$

$$r_{e7} \simeq \frac{1}{g_{m7}} = \frac{V_T}{I_{C7}} = \frac{25\text{ mV}}{10\ \mu\text{A}} = 2.5\text{ k}\Omega$$

Obviously r_{o1} and r_{o7} are very large and can be neglected. Then, the portion of $g_{m1}(v_{id}/2)$ that flows into the emitter proper of Q_7 can be found from

$$i_{e7} \simeq g_{m1}\left(\frac{v_{id}}{2}\right)\frac{R_7}{R_7 + r_{e7}}$$

$$= g_{m1}\left(\frac{v_{id}}{2}\right)\frac{20}{20+2.5} = 0.89 g_{m1}\left(\frac{v_{id}}{2}\right)$$

and the output short-circuit current i_o is

$$i_o \simeq i_{e7} = 0.89 g_{m1}(v_{id}/2)$$

Thus,

$$G_{m1} \equiv \frac{i_o}{v_{id}/2} = 0.89 g_{m1} = 0.89 \times 0.2 = 0.18 \text{ mA/V}$$

To find the voltage gain we need to determine the total resistance between the output node and ground for the circuit in Fig. 13.19(a),

$$R = R_{o9} \parallel R_{o7} \parallel (R_L/2)$$

The resistance R_{o9} is the output resistance of Q_9, which has an emitter-degeneration resistance R_9. Thus R_{o9} can be found using Eq. (8.70),

$$R_{o9} = r_{o9} + (R_9 \parallel r_{\pi 9})(1 + g_{m9} r_{o9})$$

where

$$r_{o9} = \frac{|V_{Ap}|}{I_{C9}} = \frac{20 \text{ V}}{10 \ \mu\text{A}} = 2 \text{ M}\Omega$$

$$g_{m9} = \frac{I_{C9}}{V_T} = \frac{10 \ \mu\text{A}}{25 \text{ mV}} = 0.4 \text{ mA/V}$$

$$r_{\pi 9} = \frac{\beta_P}{g_{m9}} = \frac{10}{0.4 \text{ mA/V}} = 25 \text{ k}\Omega$$

Thus

$$R_{o9} = 2 + (30 \parallel 25) \times 10^{-3} \left(1 + 0.4 \times 2 \times 10^3\right) = 12.9 \text{ M}\Omega$$

The resistance R_{o7} is the output resistance of Q_7, which has an emitter-degeneration resistance $(R_7 \parallel r_{o1}) \simeq R_7$. Thus,

$$R_{o7} = r_{o7} + (R_7 \parallel r_{\pi 7})(1 + g_{m7} r_{o7})$$

where

$$r_{o7} = \frac{V_{An}}{I_{C7}} = \frac{30 \text{ V}}{10 \ \mu\text{A}} = 3 \text{ M}\Omega$$

$$g_{m7} = \frac{I_{C7}}{V_T} = \frac{10 \ \mu\text{A}}{25 \text{ mV}} = 0.4 \text{ mA/V}$$

$$r_{\pi 7} = \frac{\beta_N}{g_{m7}} = \frac{40}{0.4} = 100 \text{ k}\Omega$$

Thus,

$$R_{o7} = 3 + (20 \parallel 100) \times 10^{-3} \left(1 + 0.4 \times 3 \times 10^3\right) = 23 \text{ M}\Omega$$

$$\frac{R_L}{2} = \frac{2 \text{ M}\Omega}{2} = 1 \text{ M}\Omega$$

Example 13.3 *continued*

The total resistance R can now be found as

$$R = 12.9 \parallel 23 \parallel 1 = 0.89\,\text{M}\Omega$$

Finally, we can find the voltage gain as

$$A_d = \frac{v_{od}/2}{v_{id}/2} = G_{m1}R$$
$$= 0.18 \times 0.89 \times 10^3 = 160\,\text{V/V}$$

13.3.3 Common-Mode Feedback to Control the DC Voltage at the Output of the Input Stage

For the cascode circuit in Fig. 13.18 to operate properly and provide high output resistance and thus high voltage gain, the cascode transistors Q_7 through Q_{10} must operate in the active mode at all times. However, relying solely on matching will not be enough to ensure that the currents supplied by Q_9 and Q_{10} are exactly equal to the currents supplied by Q_7 and Q_8. Any small mismatch ΔI between the two sets of currents will be multiplied by the large output resistance between each of the collector nodes and ground, and thus there will be large changes in the voltages v_{O1} and v_{O2}. These changes can cause one set of the current sources (i.e., Q_7-Q_8 or Q_9-Q_{10}) to saturate. We therefore need a circuit that detects the change in the dc or common-mode component V_{CM} of v_{O1} and v_{O2},

$$V_{CM} = \frac{1}{2}(v_{O1} + v_{O2}) \tag{13.73}$$

and adjusts the bias voltage on the bases of Q_7 and Q_8, V_B, to restore current equality. This negative-feedback loop should be insensitive to the differential signal components of v_{O1} and v_{O2}; otherwise it would reduce the differential gain. Thus the feedback loop should provide **common-mode feedback** (CMF).

Figure 13.20 shows the cascode circuit with the CMF circuit as a black box. The CMF circuit accepts v_{O1} and v_{O2} as inputs and provides the bias voltage V_B as output. In a particular implementation we will present shortly, the CMF circuit has the transfer characteristic

$$V_B = V_{CM} + 0.4 \tag{13.74}$$

By keeping V_B higher than V_{CM} by only 0.4 V, the CMF circuit ensures that Q_7 and Q_8 remain active (0.6 V is needed for saturation).

The nominal value of V_B is determined by the quiescent current of Q_7 through Q_{10}, the quiescent value of I_1 and I_2, and the value of R_7 and R_8. The resulting nominal value of V_B and the corresponding value of V_{CM} from Eq. (13.74) are designed to ensure that Q_9 and Q_{10} operate in the active mode. Here, it is important to recall that V_{BIAS3} is determined by the rest of the op-amp bias circuit.

To see how the CMF circuit regulates the dc voltage V_{CM}, assume that for some reason V_B is higher than it should be and as a result the currents of Q_7 and Q_8 exceed the currents

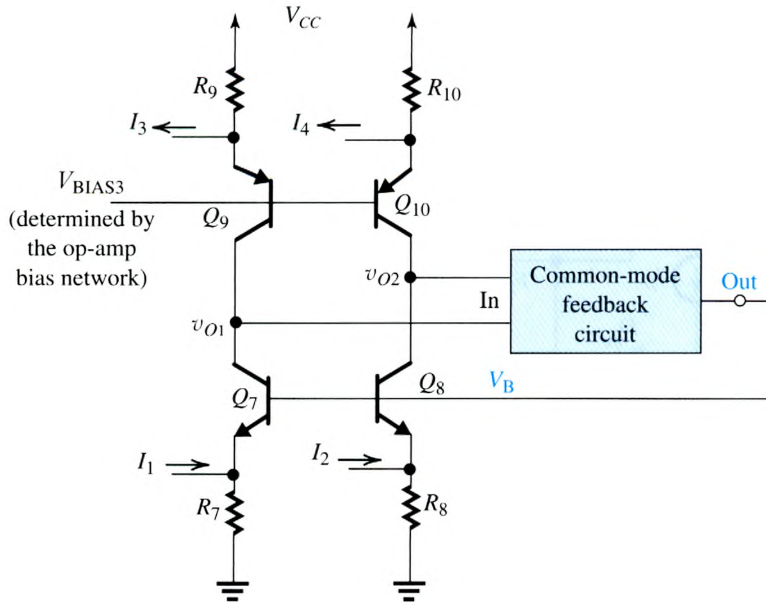


Figure 13.20 The cascode output circuit of the input stage and the CMF circuit that responds to the common-mode component $V_{CM} = \frac{1}{2}(v_{O1} + v_{O2})$ by adjusting V_B so that Q_7–Q_8 conduct equal currents to Q_9–Q_{10}, and Q_7–Q_{10} operate in the active mode.

supplied by Q_9 and Q_{10} by an increment ΔI. When multiplied by the total resistance between each of the output nodes and ground, the increment ΔI will result in a large negative voltage increment in v_{O1} and v_{O2}. The CMF circuit responds by lowering V_B to the value that restores the equality of currents. The change in V_B needed to restore equilibrium is usually small (see Example 13.4 on page 978), and according to Eq. (13.74), the corresponding change in V_{CM} will be equally small. Thus we see negative feedback in action: It minimizes the initial change and thus keeps V_{CM} nearly constant at its nominal value, which is designed to operate Q_7 through Q_{10} in the active region.

We conclude by briefly considering a possible implementation of the CMF circuit. Figure 13.21 shows the second stage of an op-amp circuit. The circuit is fed by the outputs v_{O1} and v_{O2},

$$v_{O1} = V_{CM} + v_d/2$$
$$v_{O2} = V_{CM} - v_d/2$$

In addition to amplifying the differential component of v_d, the circuit generates a dc voltage V_B,

$$V_B = V_{CM} + 0.4$$

To see how the circuit works, note that Q_{11} and Q_{12} are emitter followers that minimize the loading of the second stage on the input stage. The emitter followers deliver to the bases of the differential pair Q_{13}–Q_{14} voltages that are almost equal to v_{O1} and v_{O2} but dc shifted by $V_{EB11,12}$. Thus the voltage at the emitters of Q_{13}–Q_{14} will be

$$V_E = V_{CM} + V_{EB11,12} - V_{BE13,14}$$

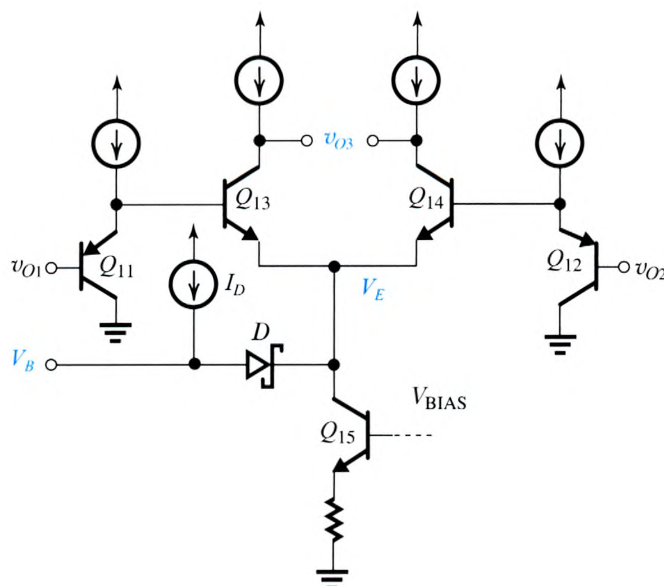


Figure 13.21 An op-amp second stage incorporating the common-mode feedback circuit for the input stage. Note that the circuit generates the voltage V_B needed to bias the cascode circuit in the first stage. Diode D is a Schottky-barrier diode, which exhibits a forward voltage drop of about 0.4 V.

which reduces to

$$V_E \simeq V_{CM}$$

The voltage V_B is simply equal to V_E plus the voltage drop of diode D. The latter is a **Schottky-barrier diode** (SBD), which features a low forward drop of about 0.4 V. Thus,

$$V_B = V_E + V_D = V_{CM} + 0.4$$

as required. This CMF circuit is effective for small differential voltages $(v_{O2} - v_{O1})$. The differential pair Q_{13-14} provides gain, so a wider swing appears at v_{O3}.

Example 13.4

Consider the operation of the circuit in Fig. 13.20. Assume that $V_{ICM} \ll 0.8$ V so the *npn* input pair (Fig. 13.18) is off. Hence $I_3 = I_4 = 0$. Also assume that only dc voltages are present and so $I_1 = I_2 = 5\ \mu$A. Each of Q_7 to Q_{10} is biased at 10 μA, $V_{CC} = 3$ V, $V_{\text{BIAS3}} = V_{CC} - 1$, $R_7 = R_8 = 20$ kΩ, and $R_9 = R_{10} = 30$ kΩ. Neglect base currents and neglect the loading effect of the CMF circuit on the output nodes of the cascode circuit. The CMF circuit provides $V_B = V_{CM} + 0.4$.

(a) Determine the nominal values of V_B and V_{CM}. Does the value of V_{CM} ensure operation in the active mode for Q_7 through Q_{10}?
(b) If the CMF circuit were not present, what would be the change in v_{O1} and v_{O2} (i.e., in V_{CM}) as a result of a current mismatch $\Delta I = 0.3\ \mu\text{A}$ between Q_7–Q_8 and Q_9–Q_{10}? Use the output resistance values found in Example 13.3.
(c) Now, if the CMF circuit is connected, what change will it cause in V_B to eliminate the current mismatch ΔI? What is the corresponding change in V_{CM} from its nominal value?

Solution

(a) We find the nominal value of V_B as follows:

$$V_B = V_{BE7} + (I_{E7} + I_1)R_7$$
$$\simeq 0.7 + (10+5) \times 10^{-3} \times 20$$
$$= 1\text{ V}$$

We can now find the nominal value of V_{CM} from

$$V_{CM} = V_B - 0.4 = 1 - 0.4 = 0.6\text{ V}$$

For Q_7–Q_8 to be active,

$$V_{CM} > V_B - 0.6 = 0.4\text{ V}$$

For Q_9–Q_{10} to be active

$$V_{CM} < V_{\text{BIAS3}} + 0.6 = V_{CC} - 1 + 0.6 = 2.6\text{ V}$$

Thus, for all four cascode transistors to operate in the active mode,

$$0.4\text{ V} < V_{CM} < 2.6\text{ V}$$

Thus the nominal value of 0.6 V ensures active-mode operation.

(b) For $I_{C9} - I_{C7} = I_{C10} - I_{C8} = \Delta I$,

$$\Delta V_{CM} = \Delta I R_{o1}$$

where R_{o1} is the output resistance between the collectors of Q_7 and Q_9 and ground,

$$R_{o1} = R_{o7} \parallel R_{o9}$$

In Example 13.3, we found that $R_{o7} = 23\,\text{M}\Omega$ and $R_{o9} = 12.9\text{ M}\Omega$; thus,

$$R_{o1} = 23 \parallel 12.9 = 8.3\,\text{M}\Omega$$

So,

$$\Delta V_{CM} = 0.3 \times 8.3 \simeq 2.5\text{ V}$$

Example 13.4 *continued*

Now if ΔV_{CM} is positive,

$$V_{CM} = 0.6 + 2.5 = 3.1 \text{ V}$$

which exceeds the 2.6 V maximum allowed value before $Q_9 - Q_{10}$ saturate. If ΔV_{CM} is negative,

$$V_{CM} = 0.6 - 2.5 = -1.9 \text{ V}$$

which is far below the +0.4 V needed to keep $Q_7 - Q_8$ in the active mode. So, in the absence of CMF, a current mismatch of $\pm 0.3\ \mu\text{A}$ would cause one set of the cascode transistors (depending on the polarity of ΔI) to saturate.

(c) With the CMF circuit in place, the feedback will adjust V_B by ΔV_B so that the currents in Q_7 and Q_8 will change by an increment equal to ΔI, thus restoring current equality. Since a change ΔV_B results in

$$\Delta I_{C7} = \Delta I_{C8} = \frac{\Delta V_B}{r_{e7} + R_7}$$

then

$$\Delta I = \frac{\Delta V_B}{r_{e7} + R_7}$$

$$\Delta V_B = \Delta I (r_{e7} + R_7)$$

$$= 0.3\ \mu\text{A}\left(\frac{25\text{ mV}}{10\ \mu\text{A}} + 20\text{ k}\Omega\right)$$

$$= 0.3 \times 22.5 = 6.75\text{ mV}$$

Correspondingly,

$$\Delta V_{CM} = \Delta V_B = 6.75\text{ mV}$$

Thus, to restore the current equality, the change required in V_B and V_{CM} is only 6.75 mV.

13.3.4 The 741 Op Amp Input Stage

The 741 op amp is an analog IC classic. It is essentially a Miller-compensated two-stage op amp with an additional output stage to provide low output resistance. Its performance is in many respects far surpassed by modern designs. However, its input stage, shown in Fig. 13.22, is different from the conventional two-stage op amp and illustrates clever techniques that are still useful for bipolar design today.

The input stage consists of transistors Q_1 through Q_4, with biasing performed by Q_5 and Q_6. Transistors Q_1 and Q_2 act as emitter followers, causing the input resistance to be high and delivering the differential input signal to the differential common-base amplifier formed

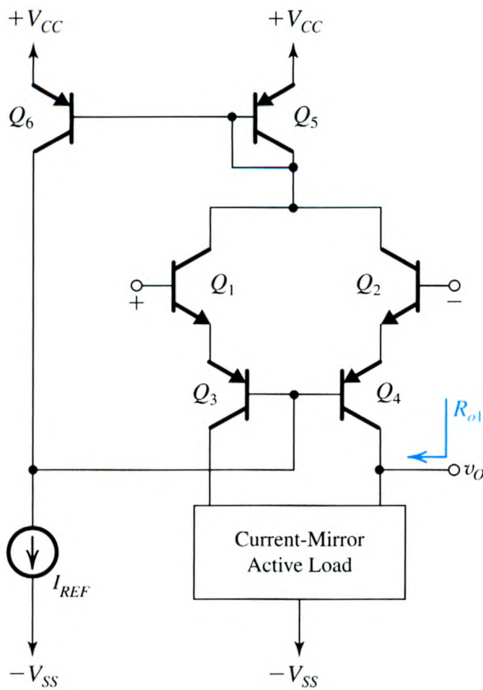


Figure 13.22 The 741 input stage.

by Q_3 and Q_4. Thus the input stage is the differential version of the common-collector, common-base configuration we discussed in Section 10.7.3.

The collectors of Q_3 and Q_4 drive a current-mirror-load circuit. As usual with current-mirror loads, it not only provides a high-resistance load for Q_4 but also converts the signal from differential to single-ended form with no loss in gain or common-mode rejection. The output of the input stage is taken single-endedly at the collector of Q_4.

As we mentioned in Section 9.6.2, every op-amp circuit includes a *level shifter* whose function is to shift the dc level of the signal so that the signal at the op-amp output can swing positive and negative. In the 741, level shifting is done in the first stage using the lateral *pnp* transistors Q_3 and Q_4. Although lateral *pnp* transistors have poor high-frequency performance, their use in the common-base configuration (which is known to have good high-frequency response) does not seriously impair the op-amp frequency response.

The use of the lateral *pnp* transistors Q_3 and Q_4 in the first stage results in an added advantage: protection of the input-stage transistors Q_1 and Q_2 against emitter–base junction breakdown. Since the emitter–base junction of an *npn* transistor breaks down at about 7 V of reverse bias (see Section 6.4.1), regular *npn* differential stages suffer such a breakdown if, say, the supply voltage (which can be as high as ±15 V) is accidentally connected between the input terminals. Lateral *pnp* transistors, however, have high emitter–base breakdown voltages (about 50 V), and because they are connected in series with Q_1 and Q_2, they provide protection of the 741 input transistors, Q_1 and Q_2.

Finally, note that except for using input buffer transistors, the 741 input stage is essentially a current-mirror-loaded differential amplifier. It is quite similar to the input stage of the CMOS amplifier in Fig. 13.1.

THE CREATOR OF THE µA741: DAVID FULLAGAR

David Fullagar was at Fairchild Semiconductor in 1967 when he designed the µA741, or simply "741," perhaps the most successful op amp ever. Educated at Cambridge, U.K., and previously employed at Ferranti, Fullagar had joined Fairchild in 1966 following Widlar's departure after the µA702 and µA709 designs. Fullagar's µA741 creation incorporated internal compensation, short-circuit protection, and a novel high-impedance input stage to resolve shortcomings of the earlier designs. Later in his career, the engineer-designer cofounded and became a vital technical contributor to Maxim Integrated Products in 1983. Texas Instruments and Analog Devices still sell updated versions of the ubiquitous 741.

DC Analysis We first perform a dc analysis of the 741 input stage, Fig. 13.23, to find the dc current in each transistor. As we will see shortly, this is a negative-feedback circuit that stabilizes the bias current of each of Q_1 to Q_4 at a value approximately equal to $I_{REF}/2$. Refer to the analysis indicated in the diagram (where β_N is assumed to be high). The sum of the collector currents of Q_1 and Q_2 ($2I$) is fed to (or sensed by) the input of the current mirror Q_5–Q_6. The output current of the mirror, which for large β_P is approximately equal to $2I$, is compared to I_{REF} at node X. The difference between the two currents ($2I/\beta_P$) establishes the base currents of Q_3 and Q_4. This is the error signal of the feedback loop. For large β_P, this current approaches zero and a node equation at X gives $2I \simeq I_{REF}$, and thus $I \simeq I_{REF}/2$.

Figure 13.23 The dc analysis of the 741 input stage.

To verify the action of the negative-feedback loop in stabilizing the value of I, assume that for some reason I increases. We see that the input current of the Q_5–Q_6 mirror increases and, correspondingly, its output current increases. Assuming that I_{REF} remains constant, the base currents in Q_3 and Q_4 must decrease to satisfy a node equation at X. This in turn decreases the value of I, which is opposite to the change we at first assumed.

EXERCISES

13.16 If $I_{REF} = 19\,\mu A$, find the value of the bias current for each of Q_1, Q_2, Q_3, and Q_4.
Ans. 9.5 μA

13.17 Determine the loop gain of the feedback loop in Fig. 13.23. Break the loop at the input of the Q_5–Q_6 mirror. Since the input resistance of the mirror is low, ground the connection of the collectors of Q_1 and Q_2. Apply an input test current i_t to the current mirror and find the feedback current that appears in the combined connection of the collectors of Q_1 and Q_2. Assume I_{REF} remains constant.
Ans. Loop gain $\simeq \beta_P$

13.18 Recalling from Chapters 2 and 9 that the input bias current of an op amp is the average of its two input currents,

$$I_B = \frac{1}{2}(I_{B1} + I_{B2})$$

and the input offset current is

$$I_{OS} = |I_{B1} - I_{B2}|$$

find I_B and I_{OS} for $I_{REF} = 19\,\mu A$ if β_1 and β_2 are nominally 200 but can deviate from nominal by as much as ±5%.
Ans. 47.5 nA; 4.75 nA

Input Common-Mode Range The **input common-mode range** is the range of input common-mode voltages over which the input stage remains in the linear active mode. Refer to Fig. 13.22. We see that in the 741 circuit the input common-mode range is determined at the upper end by saturation of Q_1 and Q_2, and at the lower end by saturation of Q_3 and Q_4, which in turn depends upon the input circuit's load.

EXERCISE

13.19 Assume that $V_{CC} = 15$ V. Show that the maximum allowable input common-mode voltage of the 741 is approximately +14.7 V. (Assume that $V_{BE} \simeq 0.6$ V and that to avoid saturation $V_{CB} \geq -0.3$ V for an *npn* transistor.)

Small-Signal Analysis Figure 13.24 shows part of the 741 input stage for the purpose of performing small-signal analysis. Note that since the collectors of Q_1 and Q_2 are connected to a constant dc voltage, they are shown grounded. Also, the constant-current biasing of the bases of Q_3 and Q_4 is equivalent to having the common-base terminal open-circuited.

The differential signal v_i applied between the input terminals effectively appears across four equal emitter resistances connected in series—those of Q_1, Q_2, Q_3, and Q_4. As a result, emitter signal currents flow as indicated in Fig. 13.24 with

$$i_e = \frac{v_i}{4r_e} \tag{13.75}$$

where r_e denotes the emitter resistance of each of Q_1 through Q_4. So

$$r_e = \frac{V_T}{I}$$

Thus the four transistors Q_1 through Q_4 supply the load circuit with a pair of complementary current signals αi_e, as indicated in Fig. 13.24.

We can find the input differential resistance of the op amp from Fig. 13.24 as

$$R_{id} = 4(\beta_N + 1)r_e \tag{13.76}$$

Assuming the input circuit is loaded by a current-mirror active load as in Fig. 13.22, the output current is

$$i_o = 2\alpha i_e \tag{13.77}$$

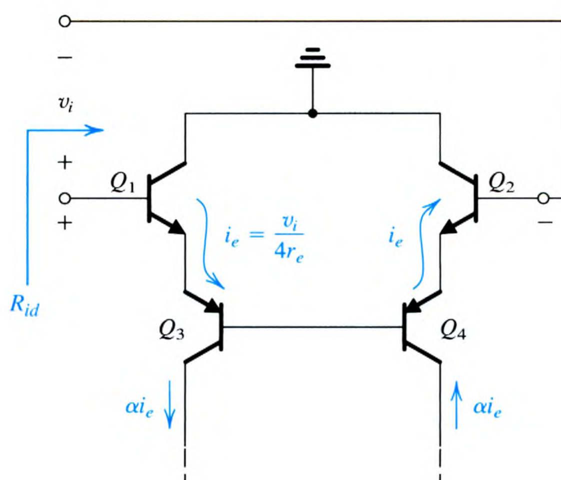


Figure 13.24 Small-signal analysis of the 741 input stage.

Combining Eqs. (13.75) and (13.77) provides the transconductance of the input stage as

$$G_{m1} \equiv \frac{i_o}{v_i} = \frac{\alpha}{2r_e} = \frac{1}{2} g_{m1} \tag{13.78}$$

where g_{m1} is the transconductance of each of the four transistors Q_1, Q_2, Q_3, and Q_4.

EXERCISE

13.20 Assuming that each of the input-stage transistors is biased at a current $I = 9.5\ \mu\text{A}$ and that $\beta_N = 200$, find r_e, g_{m1}, G_{m1}, and R_{id}.

Ans. 2.63 kΩ; 0.38 mA/V; 0.19 mA/V; 2.1 MΩ

To complete our modeling of the 741 input stage, we must find its output resistance R_{o1}. This is the resistance seen "looking back" into the output terminal of the circuit in Fig. 13.22. Thus R_{o1} is the parallel equivalent of the output resistance looking into the collector of Q_4 and the output resistance of the current-mirror active load. It's considerably easier to find the resistance looking into the collector of Q_4 if we assume that the common bases of Q_3 and Q_4 are at a *virtual ground*. This of course happens only when the input signal v_i is applied in a complementary fashion. Nevertheless, making this assumption does not result in a large error.

Assuming that the base of Q_4 is at virtual ground, the resistance we are after is R_{o4}, indicated in Fig. 13.25. This is the output resistance of a common-base transistor that has a resistance (r_e of Q_2) in its emitter. To find R_{o4} we use the following expression (Eq. 8.71):

$$R_o = r_o\left[1 + g_m(R_e \parallel r_\pi)\right] \tag{13.79}$$

where $R_e = r_e$ and $r_o = V_{Ap}/I$.

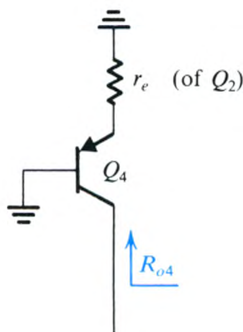


Figure 13.25 Simplified circuit for finding the output resistance R_{o4} of the first stage.

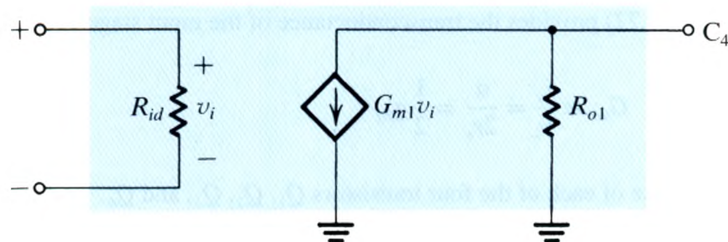


Figure 13.26 Small-signal equivalent circuit for the input stage of the 741 op amp.

The second component of the output resistance is that seen looking into the output of the current-mirror active load. In the 741, it is a bipolar current mirror with output resistance that we may also determine with Eq. (13.79).

Figure 13.26 shows the equivalent circuit that we have derived for the input stage.

EXERCISES

13.21 Find R_{o4}. Assuming the active load has a similar output resistance, find the total output resistance of the input stage, R_{o1}. Assume that $I = 9.5\ \mu\text{A}$, $V_{An} = 125$ V, $V_{Ap} = 50$ V, $\beta_N = 200$, and $\beta_P = 50$.
Ans. 10.5 MΩ; 5.3 MΩ

13.22 Use the equivalent circuit of Fig. 13.26 together with the value of G_{m1} found in Exercise 13.20 and the value of R_{o1} found in Exercise 13.21 to determine the open-circuit voltage gain of the 741 input stage.
Ans. $|A_{vo}| = G_{m1}R_{o1} = 1007$ V/V

Example 13.5

It is required to find the CMRR of the 741 input stage. Assume that the circuit is balanced except for mismatches in the current-mirror load that result in an error ϵ_m in the mirror's current-transfer ratio; that is, the ratio becomes $(1-\epsilon_m)$.

Solution

In Section 9.5.5, we analyzed the common-mode operation of the current-mirror-loaded differential amplifier and derived an expression for its CMRR. The situation in the 741 input stage, however,

differs substantially because of the feedback loop that regulates the bias current. Since this feedback loop is sensitive to the common-mode signal, as we will see shortly, the loop operates to reduce the common-mode gain and, correspondingly, to increase the CMRR. Hence, its action is referred to as **common-mode feedback.**

Figure 13.27 shows the 741 input stage with a common-mode signal v_{icm} applied to both input terminals. We have assumed that as a result of v_{icm}, a signal current i flows as shown. Since the stage is balanced, both sides carry the same current i.

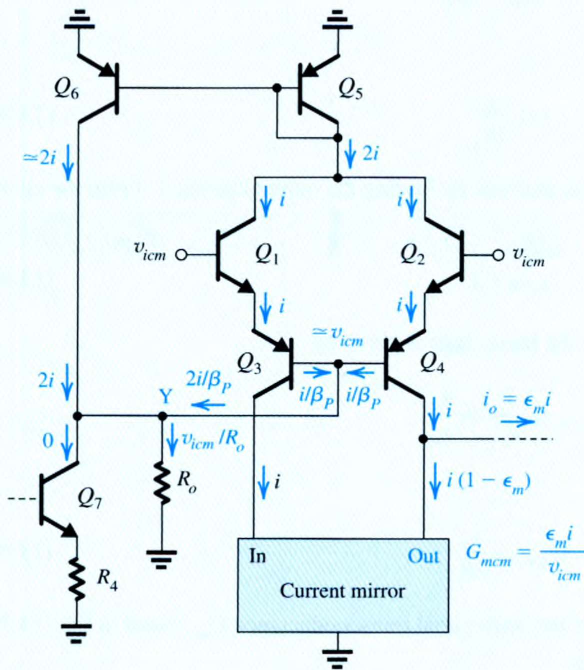


Figure 13.27 Analysis of the common-mode gain of the 741 input stage. Note that $R_o = R_{o6} \| R_{o7}$ has been "pulled out" and shown separately, leaving behind ideal current sources Q_6 and Q_7.

Our objective now is to determine how i relates to v_{icm}. Toward that end, notice that for common-mode inputs, both sides of the differential amplifier, that is, $Q_{1,3}$ and $Q_{2,4}$, act as followers, delivering a signal almost equal to v_{icm} to the common-base node of Q_3 and Q_4. Now, this node Y is connected to the collectors of two current sources, Q_6 and Q_7. Denoting the total resistance between node Y and ground R_o, we write

$$R_o = R_{o6} \| R_{o7} \tag{13.80}$$

Example 13.5 *continued*

In Fig. 13.27 we have "pulled R_o out," thus leaving behind ideal current sources Q_6 and Q_7. Since the current in Q_7 is constant, we show Q_7 in Fig. 13.27 as having a zero incremental current. Transistor Q_6, on the other hand, provides a current approximately equal to the one fed into Q_5, which is $2i$. This is the feedback current. Since Q_5 senses the *sum* of the currents in the two sides of the differential amplifier, the feedback loop operates only on the common-mode signal and is insensitive to any difference signal.

Proceeding with the analysis, we now can write a node equation at Y,

$$2i + \frac{2i}{\beta_P} = \frac{v_{icm}}{R_o} \tag{13.81}$$

Assuming $\beta_P \gg 1$, this equation simplifies to

$$i \simeq \frac{v_{icm}}{2R_o} \tag{13.82}$$

Having determined i, we can now complete our analysis by finding the output current i_o. From the circuit in Fig. 13.27, we see that

$$i_o = \epsilon_m i \tag{13.83}$$

Thus the common-mode transconductance of the input stage is given by

$$G_{mcm} \equiv \frac{i_o}{v_{icm}} = \frac{\epsilon_m i}{v_{icm}}$$

Substituting for i from Eq. (13.82) gives

$$G_{mcm} = \frac{\epsilon_m}{2R_o} \tag{13.84}$$

Finally, we can find the CMRR as the ratio of the differential transconductance G_{m1} found in Eq. (13.78) and the common-mode transconductance G_{mcm},

$$\text{CMRR} \equiv \frac{G_{m1}}{G_{mcm}} = g_{m1}R_o/\epsilon_m \tag{13.85}$$

where g_{m1} is the transconductance of Q_1. Now, substituting for R_o from Eq. (13.80), we obtain

$$\text{CMRR} = g_{m1}(R_{o6} \parallel R_{o7})/\epsilon_m \tag{13.86}$$

Before leaving this example, we notice that if the feedback were not present, the $2i$ term in Eq. (13.81) would be absent and the current i would become $\beta_P(v_{icm}/2R_o)$, which is β_P times higher than that when feedback is present. In other words, common-mode feedback reduces i, hence the common-mode transconductance and the common-mode gain, by a factor β_P. It can be shown that β_P is the magnitude of the loop gain. (See Exercise 13.17.)

EXERCISES

13.23 Refer to Fig. 13.27 and assume that the bases of Q_6 and Q_7 are at approximately constant voltages (signal ground). Find R_{o6}, R_{o7}, and hence R_o. Use $V_A = 125$ V for *npn* and 50 V for *pnp* transistors. Assume the bias current $I_{C6,7} = 19\ \mu$A and $R_4 = 5$ kΩ.
Ans. $R_{o6} = 2.63\,\text{M}\Omega$; $R_{o7} = 31.6\,\text{M}\Omega$; $R_o = 2.43\,\text{M}\Omega$

13.24 Use the result of Example 13.5 and Exercises 13.20 and 13.23, and assume $\epsilon_M = 0.0055$ to determine G_{mcm} and CMRR of the 741 input stage. What would the CMRR be if the common-mode feedback were not present? Assume $\beta_P = 50$.
Ans. $G_{mcm} = 1.13 \times 10^{-6}$ mA/V; CMRR $= 1.68 \times 10^5$ or 104.5 dB; without common-mode feedback, CMRR = 70.5 dB

13.3.5 Output-Stage Design for Near Rail-to-Rail Output Swing

As we mentioned earlier, modern low-voltage bipolar op amps cannot afford to use the classical emitter-follower-based class AB output stage in Fig. 13.15, which would consume too much of the power-supply voltage. Instead, we use a complementary pair of common-emitter transistors, as shown in Fig. 13.28. The output transistors Q_P and Q_N are operated in a class AB fashion. Typically, i_L can be 10 mA or higher and is determined by v_O and R_L. For $i_L = 0$, $i_P = i_N = I_Q$, where the quiescent current I_Q is normally a fraction of a milliamp.

The output stage in Fig. 13.28 is driven by two *separate but equal signals*, v_{BP} and v_{BN}. When v_{BP} and v_{BN} are high, Q_N supplies the load current in the direction opposite to that shown[8] and the output voltage v_O can swing to within 0.1 V or so of ground. In the meantime, Q_P is inactive. Nevertheless, in order to minimize crossover distortion, Q_P is prevented from turning off and is forced (as we will show shortly) to conduct a minimum current of about $I_Q/2$.

The opposite happens when v_{BP} and v_{BN} are low: Q_P supplies the load current i_L in the direction indicated, and v_O can go up as high as $V_{CC} - 0.1$ V. In the meantime, Q_N is inactive but is prevented from turning off and forced to conduct a minimum current of about $I_Q/2$.

From the description above, we see that v_O can swing to within 0.1 V of each of the supply rails. This near rail-to-rail operation is the major advantage of this CE output stage. Its disadvantage is the relatively high output resistance. However, given that the op amp will almost always be used with a negative-feedback loop, the closed-loop output resistance can still be very low.

A Buffer/Driver Stage The output transistors can be called on to supply large currents (e.g., in the 10 mA to 15 mA range). When this happens, the base currents of Q_P and Q_N can be substantial, especially since these output BJTs typically have modest current gains (e.g., $\beta_P \simeq 10$ and $\beta_N \simeq 40$). Such large currents cannot usually be supplied directly by the amplifier stage preceding the output stage. Rather, we often need a buffer/driver stage, as shown in Fig. 13.29. Here we use an emitter follower Q_3 to drive Q_N. However, because of

[8] For this to happen, either R_L is returned to the positive supply (rather than ground) or R_L is capacitively coupled to the amplifier output.

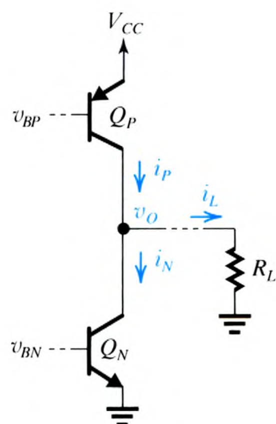


Figure 13.28 In order to provide v_O that can swing to within 0.1 V of V_{CC} and ground, a near rail-to-rail operation, the output stage uses common-emitter transistors. Note that the driving signals v_{BP} and v_{BN} are separate but identical.

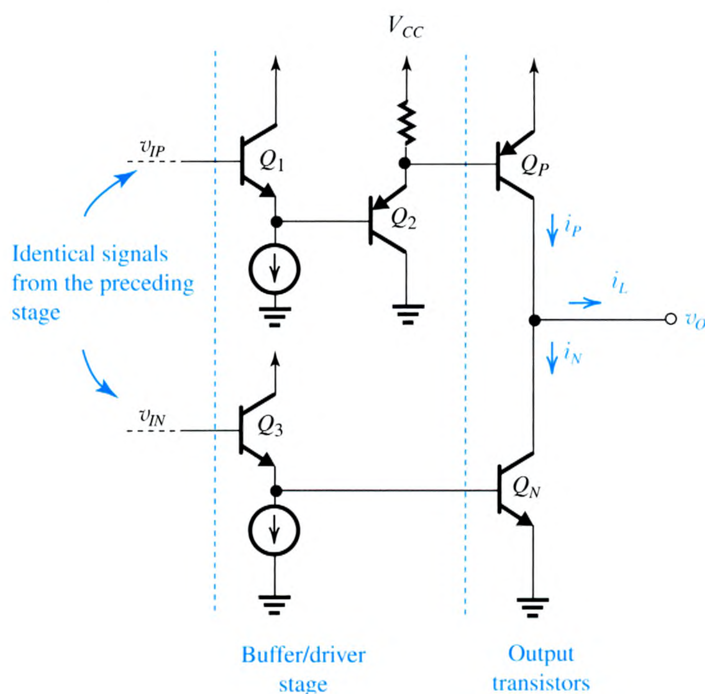


Figure 13.29 The output stage that is operated as class AB needs emitter-follower buffers/drivers to reduce the loading on the preceding stage and to provide the current gain necessary to drive Q_P and Q_N.

the low β_P, we use a double buffer consisting of complementary emitter followers Q_1 and Q_2 to drive Q_P. The driver stage is fed by two separate but identical signals v_{IP} and v_{IN} that come from the preceding amplifier stage (which is usually the second stage) in the op-amp circuit.[9]

[9] An interesting approach for generating two identical outputs in the second stage is used in the NE5234 (see Gray et al., 2009).

EXERCISE

13.25 (a) For the circuit in Fig. 13.29, find the current gain from each of the v_{IP} and v_{IN} terminals to the output in terms of β_P and β_N.

(b) For $i_L = \pm 10$ mA, how much signal current is needed at the v_{IP} and v_{IN} inputs?

Ans. (a) $\beta_N\beta_P^2$, β_N^2; (b) 2.5 μA, 6.25 μA

Establishing I_Q and Maintaining a Minimum Current in the Inactive Transistor

Next we'll consider the circuit for establishing the quiescent current I_Q in Q_N and Q_P and for maintaining a minimum current of $I_Q/2$ in the inactive output transistor. Figure 13.30 shows a fuller version of the output stage. In addition to the output transistors Q_P–Q_N and the buffer/driver stage, which we have already discussed, the circuit includes two circuit blocks whose operation we will now explain.

The first is the circuit composed of the differential pair Q_6–Q_7 and associated transistors Q_4 and Q_5, and resistors R_4 and R_5. This circuit measures the currents in the output transistors, i_P and i_N, and arranges for the current I to divide between Q_6 and Q_7 according to the ratio i_N/i_P, and provides a related output voltage v_E. Specifically, it can be shown (Problems 13.68 and 13.69) that

$$i_{C6} = I\,\frac{i_N}{i_P + i_N} \tag{13.87}$$

$$i_{C7} = I\,\frac{i_P}{i_P + i_N} \tag{13.88}$$

$$v_E = V_T \ln\left[\frac{i_N i_P}{i_N + i_P}\,\frac{I}{I_{SN} I_{S7}}\right] \tag{13.89}$$

where I_{SN} and I_{S7} are the saturation currents of Q_N and Q_7, respectively. Observe that for $i_P \gg i_N$, $i_{C6} \simeq 0$ and $i_{C7} \simeq I$. Thus Q_6 turns off and Q_7 conducts all of I. The emitter voltage v_E becomes

$$v_E \simeq V_T \ln\left(\frac{i_N}{I_{SN}}\right) + V_T \ln\left(\frac{I}{I_{S7}}\right)$$

Thus,

$$v_E = V_T \ln\left(\frac{i_N}{I_{SN}}\right) + V_{EB7} \tag{13.90}$$

This equation simply states that $v_E = v_{BEN} + V_{EB7}$, which could have been found directly from the circuit diagram in Fig. 13.30. The important point to note, however, is that since V_{EB7} is a constant, v_E is determined by the current i_N in the inactive transistor, Q_N. In the other extreme case of $i_N \gg i_P$, $i_{C6} \simeq I$, $i_{C7} \simeq 0$; thus Q_7 turns off and Q_6 conducts all of I. In this case we can

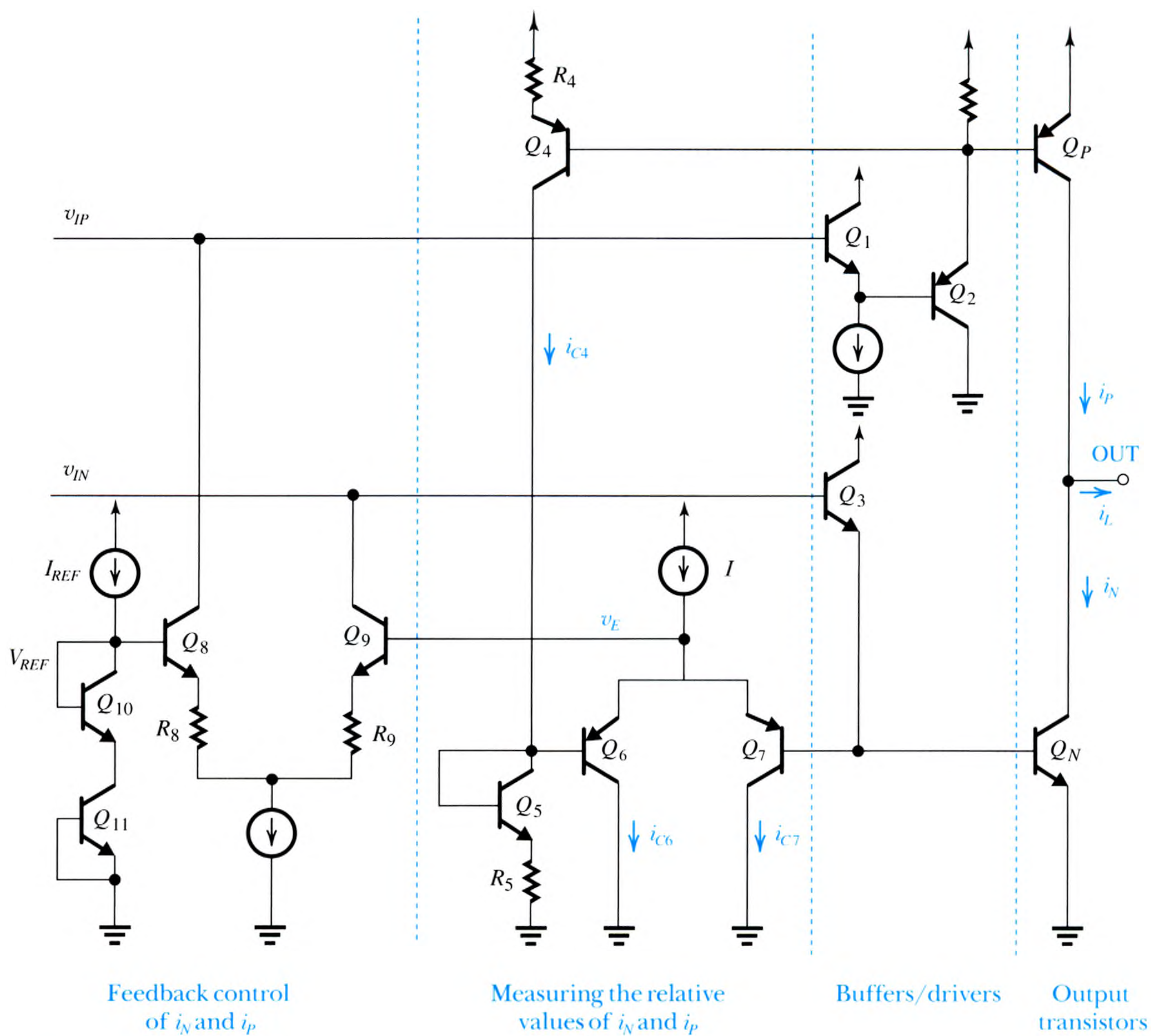


Figure 13.30 A more complete version of the output stage showing the circuits that establish the quiescent current in Q_P and Q_N. As well, this circuit forces a minimum current of $(I_Q/2)$ to follow in the inactive output transistor, preventing the transistor from turning off and thus minimizing crossover distortion.

use Eq. (13.89) to show that

$$v_E = V_T \ln\left(\frac{i_P}{I_{SN}}\right) + V_{EB6} \tag{13.91}$$

So, here too, since V_{EB6} is a constant, v_E is determined by the current in the inactive transistor, Q_P.

The second circuit block is a differential amplifier composed of Q_8–Q_9 with their emitter-degeneration resistors R_8, R_9. The voltage v_E generated by the measuring circuit is fed to one input of the differential amplifier, and the other input is fed with a reference voltage V_{REF} generated by passing a reference current I_{REF} through the series connection

of diode-connected transistors Q_{10} and Q_{11}. This differential amplifier takes part in a negative-feedback loop that uses the value of v_E to control the currents i_P and i_N through the nodes v_{IP} and v_{IN}. The objective of the feedback control is to set the current in the inactive output transistor to a minimum value. To see how the feedback operates, consider the case when $i_P \gg i_N$, and thus Q_N is the inactive transistor. In this case, Q_6 turns off, Q_7 conducts all of I, and v_E is given by Eq. (13.90). Now, if for some reason i_N falls below its minimum intended value, v_E decreases, causing i_{C9} to decrease. This in turn will cause the node v_{IN} to rise and the voltage at the base of Q_N will eventually rise, thus increasing i_N to its intended value.

Analytically, we can obtain a relationship between i_N and i_P as follows. Assume that the loop gain of the feedback loop that is anchored by the differential amplifier Q_8–Q_9 is high enough to force the two input terminals to the same voltage, that is,

$$v_E = V_{\text{REF}} = V_T \ln \frac{I_{\text{REF}}}{I_{S10}} + V_T \ln \frac{I_{\text{REF}}}{I_{S11}}$$

Substituting for v_E from Eq. (13.89) results in

$$\frac{i_N i_P}{i_N + i_P} = \left(\frac{I_{\text{REF}}^2}{I}\right)\left(\frac{I_{SN}}{I_{S10}}\right)\left(\frac{I_{S7}}{I_{S11}}\right) \tag{13.92}$$

Observe that the quantity on the right-hand side is a constant. In the quiescent case, $i_N = i_P = I_Q$, Eq. (13.92) yields

$$I_Q = 2\left(\frac{I_{\text{REF}}^2}{I}\right)\left(\frac{I_{SN}}{I_{S10}}\right)\left(\frac{I_{S7}}{I_{S11}}\right) \tag{13.93}$$

Thus, the constant on the right-hand side of Eq. (13.92) is $I_Q/2$, and we can rewrite Eq. (13.92) as

$$\frac{i_N i_P}{i_N + i_P} = \frac{1}{2} I_Q \tag{13.94}$$

Equation (13.94) clearly shows that for $i_N \gg i_P$, $i_P \simeq \frac{1}{2} I_Q$, and that for $i_P \gg i_N$, $i_N \simeq \frac{1}{2} I_Q$. Thus the circuit not only establishes the quiescent current I_Q (Eq. 13.93), but also sets the minimum current in the inactive output transistor at $\frac{1}{2} I_Q$.

EXERCISE

D13.26 For the circuit in Fig. 13.30, determine the value that I_{REF} should have so that Q_N and Q_P have a quiescent current $I_Q = 0.4$ mA. Assume that the transistor areas are scaled so that $I_{SN}/I_{S10} = 10$ and $I_{S7}/I_{S11} = 2$. Let $I = 10\ \mu$A. Also, if i_L in the direction out of the amplifier is 10 mA, find i_P and i_N.
Ans. $I_{\text{REF}} = 10\ \mu$A; $i_P \simeq 10.2$ mA, $i_N \simeq 0.2$ mA

Summary

- Most CMOS op amps are designed to operate as part of a VLSI circuit and thus are required to drive only small capacitive loads. Therefore, most do not have a low-output-resistance stage.
- We studied two approaches for the design of CMOS op amps: a two-stage configuration and a single-stage topology using the folded-cascode circuit.
- In the two-stage CMOS op amp, approximately equal gains are realized in the two stages.
- Device mismatch together with the lower transconductance of the input stage result in a larger input offset voltage for CMOS op amps than for bipolar units.
- Miller compensation is used in the two-stage CMOS op amp locating the dominant pole at a very low frequency using a relatively small capacitor. A series resistor is required to place the transmission zero on the negative real axis so as to increase the phase margin.
- The output capacitance, including the load, of the two-stage op amp determines the frequency of the second, nondominant pole. Therefore, increasing the load capacitance decreases the phase margin.
- The slew rate of a two-stage op amp is determined by the first-stage bias current and the frequency-compensation capacitor.
- Use of the cascode configuration increases the gain of a CMOS amplifier stage by about two orders of magnitude, thus making possible a single-stage op amp.
- The dominant pole of the folded-cascode op amp is determined by the total capacitance at the output node, C_L. Increasing C_L improves the phase margin at the expense of reducing the bandwidth.
- By using two complementary input differential pairs in parallel, the input common-mode range can be extended to equal the entire power-supply voltage, providing so-called rail-to-rail operation at the input.
- The output voltage swing of the folded-cascode op amp can be extended by utilizing a wide-swing current mirror in place of the cascode mirror.
- The rail-to-rail input common-mode range is achieved in BJT op amps by using resistive loads (instead of current-mirror loads) for the input differential pair as well as utilizing two complementary differential amplifiers in parallel.
- To increase the gain of the input stage above that achieved with resistive loads, we use the folded-cascode configuration.
- To regulate the dc bias voltages at the outputs of the differential folded-cascode stage so as to maintain active-mode operation at all times, we use common-mode feedback.
- To obtain low input offset voltage and current, and high CMRR, the 741 input stage is designed to be perfectly balanced. The CMRR is increased by common-mode feedback, which also stabilizes the dc operating point.
- To obtain high input resistance and low input bias current, the input stage of the 741 is operated at a very low current level.
- The output stage of a low-voltage BJT op amp uses a complementary pair of common-emitter transistors. This allows v_O to swing to within 0.1 V or so from each of the supply rails. The disadvantage is a high open-loop output resistance. This, however, is substantially reduced when negative feedback is applied around the op amp.
- Modern output stages of BJT op amps operate in the class AB mode and utilize interesting feedback techniques to set the quiescent current as well as to ensure that the inactive output transistor does not turn off, a precaution that avoids increases in crossover distortion.

PROBLEMS

Computer Simulation Problems

SIM Problems identified by this icon are intended to demonstrate the value of using SPICE simulation to verify hand analysis and design, and to investigate important issues such as allowable signal swing and amplifier nonlinear distortion. Instructions to assist in setting up simulations for all the indicated problems can be found in the corresponding

---

problems with blue numbers are considered essential; * = difficult problem; ** = more difficult; *** = very challenging
SIM = simulation; **D** = design problem; ● = see related video example

files on the website. Note that if a particular parameter value is not specified in the problem statement, you are to make a reasonable assumption.

Section 13.1: The Two-Stage CMOS Op Amp

13.1 A particular design of the two-stage CMOS operational amplifier of Fig. 13.1 utilizes ±1.5-V power supplies. All transistors are operated at overdrive voltages of 0.2-V magnitude. The process technology provides devices with $V_{tn} = |V_{tp}| = 0.4$ V. Find the input common-mode range and the range allowed for v_O.

13.2 The CMOS op amp of Fig. 13.1 is fabricated in the 0.18-μm process of Appendix K for which $V'_{An} = 5$ V/μm and $|V'_{Ap}| = 6$ V/μm. Find A_1, A_2, and A_v if all devices are 0.3 μm long, Q_1 and Q_2 are operated at overdrive voltages of 0.15-V magnitude, and Q_6 is operated at $V_{OV} = 0.2$ V. Also, determine the op-amp output resistance obtained when the second stage is biased at 0.3 mA. What do you expect the output resistance of a unity-gain voltage amplifier to be, using this op amp?

D 13.3 The CMOS op amp of Fig. 13.1 is fabricated in a process for which $|V'_A|$ for all devices is 6 V/μm. If all transistors have $L = 0.5$ μm and are operated at equal overdrive voltages, find the magnitude of the overdrive voltage required to obtain a dc open-loop gain of 500 V/V.

13.4 Consider the circuit in Fig. 13.1 with the device geometries shown at the bottom of this page. Let $I_{REF} = 40$ μA, $|V_t|$ for all devices = 0.45 V, $\mu_n C_{ox} = 390$ μA/V^2, $\mu_p C_{ox} = 70$ μA/V^2, $|V_A|$ for all devices = 5 V, $V_{DD} = V_{SS} = 1$ V. Determine the size of Q_6, W/L, that will ensure that the op amp will not have a systematic offset voltage. Then, for all devices, evaluate I_D, $|V_{OV}|$, $|V_{GS}|$, g_m, and r_o. Provide your results in a table. Also find A_1, A_2, the power consumption, the dc open-loop voltage gain, the input common-mode range, and the output voltage range. Neglect the effect of V_A on the bias currents.

D 13.5 Design the two-stage CMOS op amp in Fig. 13.1 to provide a CMRR of about 72 dB. If all the transistors are operated at equal overdrive voltages of 0.15 V and have equal channel lengths, find the minimum required channel length. For this technology, $|V'_A| = 15$ V/μm. What is the dc gain realized?

13.6 A two-stage CMOS op amp has $G_{m1} = 0.8$ mA/V, $G_{m2} = 2.4$ mA/V, $C_1 = 0.1$ pF, and $C_2 = 0.8$ pF. Find the value of C_C that will provide a unity-gain frequency of 160 MHz. Also, determine the values of f_{P2} and f_Z.

13.7 For the CMOS amplifier in Fig. 13.1, whose equivalent circuit is shown in Fig. 13.2, let $G_{m1} = 1$ mA/V, $R_1 = 40$ kΩ, $C_1 = 0.1$ pF, $G_{m2} = 1.5$ mA/V, $R_2 = 20$ kΩ, and $C_2 = 1$ pF.

(a) Find the dc gain.
(b) Without C_C connected, find the frequencies of the two poles in radians per second and sketch a Bode plot for the gain magnitude.
(c) With C_C connected, find ω_{P2}. Then find the value of C_C that will result in a unity-gain frequency ω_t at least two octaves below ω_{P2}. For this value of C_C, find ω_{P1} and ω_Z and sketch a Bode plot for the gain magnitude.

13.8 A CMOS op amp with the topology in Fig. 13.1 has $g_{m1} = g_{m2} = 1$ mA/V, $g_{m6} = 3$ mA/V, the total capacitance between node D_2 and ground is 0.2 pF, and the total capacitance between the output node and ground is 3 pF. Find the value of C_C that results in $f_t = 50$ MHz, and verify that f_t is lower than f_Z and f_{P2}. Estimate the bias drain currents in $Q_{1,2}$ and Q_6 assuming $|V_{OV}| = 0.2$ V for all transistors.

13.9 A particular design of the two-stage CMOS op amp of Fig. 13.1 has $G_{m1} = 1.5$ mA/V and $G_{m2} = 2$ mA/V. The total capacitance at the output node is 1 pF and the dc gain is 70 dB. While utilizing a Miller compensation capacitor C_C without a series resistance R, the amplifier is made to have a uniform −20-dB/decade gain rolloff with a unity-gain frequency f_t of 100 MHz.

(a) What must the value of C_C be?
(b) What do you estimate the frequencies of the poles, f_{P1} and f_{P2}, and of the right-half-plane zero, f_Z, to be?
(c) What is the phase margin obtained?
(d) To increase the phase margin, a resistance R is connected in series with C_C. What is the value of R that results in $f_Z = \infty$, and what is the resulting phase margin?

Transistor	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8
W/L (μm/μm)	36	36	6	6	30	W	45	6

(e) If R is increased further, until it moves the zero into the left half-plane and thus introduces phase lead, what value of R is needed to obtain a phase margin of 85°?

13.10 A two-stage CMOS op amp modeled by Fig. 13.2 is internally compensated for 70° phase margin at a unity-gain frequency $f_t = 150$ MHz with $C_C = 0.8$ pF and $C_2 = 0.5$ pF. If an additional load capacitance of 2 pF is added to the output in parallel with C_2, what value of additional external compensation capacitance must be added in parallel with C_C to maintain the same phase margin? What is the resulting unity-gain frequency?

D 13.11 A particular implementation of the CMOS amplifier of Figs. 13.1 and 13.2 provides $G_{m1} = 0.3$ mA/V, $G_{m2} = 0.6$ mA/V, $r_{o2} = r_{o4} = 222$ kΩ, $r_{o6} = r_{o7} = 111$ kΩ, and $C_2 = 1$ pF.

(a) Find the dc gain.
(b) Find the frequency of the second pole, f_{P2}.
(c) Find the value of the resistance R that, when placed in series with C_C, causes the transmission zero to be located at $s = \infty$.
(d) With R in place, as in (c), find the value of C_C that results in the highest possible value of f_t while providing a phase margin of 80° in a unity-gain configuration. What value of f_t is realized? What is the corresponding frequency of the dominant pole?
(e) To what value should C_C be changed to double the value of f_t? At the new value of f_t, what is the phase shift introduced by the second pole? To reduce this excess phase shift to 10° and thus obtain an 80° phase margin, as before, what value should R be changed to?

13.12 A two-stage CMOS op amp has each of its first-stage transistors Q_1 and Q_2 operating at an overdrive voltage of 0.2 V. The op amp has a uniform −20-dB/decade frequency response with a unity-gain frequency of 50 MHz. What do you expect the slew rate of this amplifier to be? If each of Q_1 and Q_2 is biased at 50 μA, what must the value of C_C be?

13.13 An op amp with $f_t = 1$ MHz has a phase margin of 80°. If the excess phase shift is due to a second single pole, what is the frequency of this pole?

13.14 An op amp with $f_t = 1$ MHz has a phase margin of 80°. If the op amp has nearly coincident second and third poles, what is their frequency?

D *13.15 For an op amp with 100 dB dc gain whose second pole is at 5 MHz, what dominant-pole frequency is required for 85° phase margin with a closed-loop gain of 100? What is the resulting phase margin at a closed-loop gain of 50?

13.16 An internally compensated op amp having an f_t of 5 MHz and dc gain of 10^6 utilizes Miller compensation around an inverting amplifier stage with a gain of −1000. If space exists for at most a 50-pF capacitor, what resistance level must be reached at the input of the Miller amplifier for compensation to be possible?

13.17 Consider the integrator op-amp model shown in Fig. 13.3. For $G_{m1} = 2$ mA/V, $C_C = 100$ pF, sketch and label a Bode plot for the magnitude of the open-loop gain. Compare this to the Bode plot of a lowpass STC circuit having the same unity-gain frequency and a dc gain of 92 dB.

13.18 For an amplifier with a slew rate of 10 V/μs, what is the full-power bandwidth (see Example 2.9) for outputs of ±5 V? What unity-gain bandwidth, ω_t, would you expect if the topology is a two-stage CMOS op amp with $|V_{OV,1}| = 0.2$ V?

D 13.19 A two-stage CMOS op amp similar to that in Fig. 13.1 is found to have a capacitance between the output node and ground of 0.7 pF. If it is desired to have a unity-gain bandwidth f_t of 100 MHz with a phase margin of 72° in a unity-gain configuration, what must g_{m6} be set to? If $|V_{OV,6}| = 0.2$ V, what is the bias current I_{D6}? Assume that a resistance R is connected in series with the frequency-compensation capacitor C_C and adjusted to place the transmission zero at infinity. What value should R have? If the first stage is operated at $|V_{OV,1}| = 0.15$ V, what is the value of slew rate obtained? If the first-stage bias current $I = 100$ μA, what is the required value of C_C?

D 13.20 A CMOS op amp with the topology shown in Fig. 13.1 is designed to provide $G_{m1} = 0.5$ mA/V and $G_{m2} = 2$ mA.

(a) Find the value of C_C that results in $f_t = 80$ MHz.
(b) What is the maximum value that C_2 can have while achieving a 70° phase margin?

D 13.21 A CMOS op amp with the topology shown in Fig. 13.1 but with a resistance R included in series with C_C is designed to provide $G_{m1} = 0.8$ mA/V and $G_{m2} = 2$ mA/V.

(a) Find the value of C_C that results in $f_t = 100$ MHz.
(b) For $R = 500\,\Omega$, what is the maximum allowed value of C_2 for which a phase margin of at least 60° is obtained in a unity-gain configuration?

13.22 Recall that when an amplifier having a dominant pole at ω_{P1} and gain A_0 is used in a feedback amplifier with feedback factor β, the bandwidth of the closed-loop amplifier is $\omega_{P1}(1 + A_0\beta)$. When an op amp having a uniform –20-dB/decade frequency response with unity-gain frequency f_t is used in a noninverting configuration with a dc gain ideally $1/\beta$, show that the bandwidth of the closed-loop amplifier is approximately βf_t.

13.23 A two-stage CMOS op amp resembling that in Fig. 13.1 is found to have a slew rate of 60 V/μs and a unity-gain bandwidth f_t of 60 MHz.

(a) Estimate the value of the overdrive voltage at which the input-stage transistors are operating.
(b) If the first-stage bias current $I = 120\ \mu$A, what value of C_C must be used?
(c) For a process for which $\mu_p C_{ox} = 60\ \mu\text{A/V}^2$, what W/L ratio applies for Q_1 and Q_2?

D 13.24 Sketch the circuit of a two-stage CMOS amplifier having the structure of Fig. 13.1 but utilizing NMOS transistors in the input stage (i.e., Q_1 and, Q_2).

D 13.25 (a) Show that the PSRR$^-$ of a CMOS two-stage op amp for which all transistors have the same channel length and are operated at equal $|V_{OV}|$ is given by

$$\text{PSRR}^- = 2\left|\frac{V_A}{V_{OV}}\right|^2$$

(b) For $|V_{OV}| = 0.15$ V, what is the minimum channel length required to obtain a PSRR$^-$ of 72 dB? For the technology available, $|V_A'| = 15$ V/μm.

Section 13.2: The Folded-Cascode CMOS Op Amp

D 13.26 The op-amp circuit of Fig. 13.9 is operated from ±1-V power supplies. If the power dissipated in the circuit is to be limited to 1 mW, find the maximum value of I_B allowed. If this value is used, and each of Q_1 and Q_2 is to be biased at a current four times that used for each of Q_3 and Q_4, find the value of I, $I_{D1,2}$, and $I_{D3,4}$.

D 13.27 For the folded-cascode op amp in Fig. 13.10 utilizing power supplies of ±1 V, find the values of $V_{\text{BIAS1}}, V_{\text{BIAS2}}$, and V_{BIAS3} to maximize the allowable range of V_{ICM} and v_O. Assume that all transistors are operated at equal overdrive voltages of 0.15 V. Assume $|V_t|$ for all devices is 0.4 V. Specify the maximum range of V_{ICM} and of v_O.

D 13.28 For the folded-cascode op-amp circuit of Figs. 13.9 and 13.10 with bias currents $I = 400\ \mu$A and $I_B = 250\ \mu$A, and with all transistors operated at overdrive voltages of 0.2 V, find the W/L ratios for all devices. Assume the 0.13-μm CMOS technology of Appendix K having $k_n' = 511\ \mu\text{A/V}^2$ and $k_p' = 128\ \mu\text{A/V}^2$.

13.29 Consider a design of the cascode op amp of Fig. 13.10 for which $I = 200\ \mu$A and $I_B = 150\ \mu$A. Assume that all transistors are operated at $|V_{OV}| = 0.2$ V and that for all devices, $|V_A| = 5$ V. Find G_m, R_o, and A_v. Also, if the op amp is connected in the feedback configuration shown in Fig. P13.29, find the voltage gain and output resistance of the closed-loop amplifier.

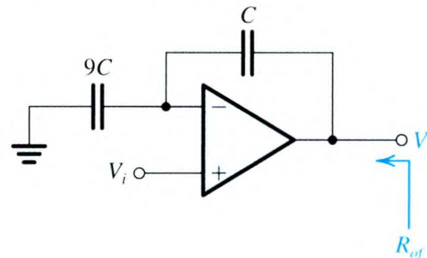


Figure P13.29

D 13.30 Consider the folded-cascode op amp of Fig. 13.9 when loaded with a 10-pF capacitance. What should the bias current I_B be to obtain a slew rate of at least 10 V/μs? If the input-stage transistors are biased at a current three times that at which each of Q_3 and Q_4 is biased, find the value of I. If the input-stage transistors are operated at overdrive voltages of 0.15 V, what is the unity-gain bandwidth realized? If the two nondominant poles have the same frequency of 50 MHz, what is the phase margin obtained in a unity-gain configuration? If it is required to have a phase margin of 75°, what must f_t be reduced to? By what amount should C_L be increased? What is the new value of SR?

13.31 For a particular design of the folded-cascode op amp in Fig. 13.9, $I < I_B$. What slew rate is obtained?

13.32 A folded-cascode amplifier is connected to its capacitive load C_L through a series resistor R_S as shown in Fig. P13.32. Find the frequency responses V_o/V_{id} and V_f/V_{id}. Show that V_f/V_{id} has a zero at $f_z = 1/2\pi R_S C_L$. How may this technique be used in frequency compensation?

D *13.33 Design the folded-cascode circuit of Fig. 13.10 to provide voltage gain of 80 dB and a unity-gain frequency of

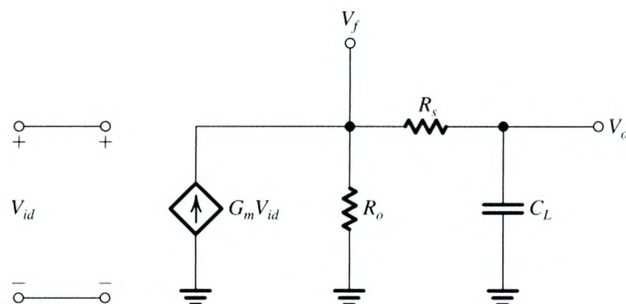


Figure P13.32

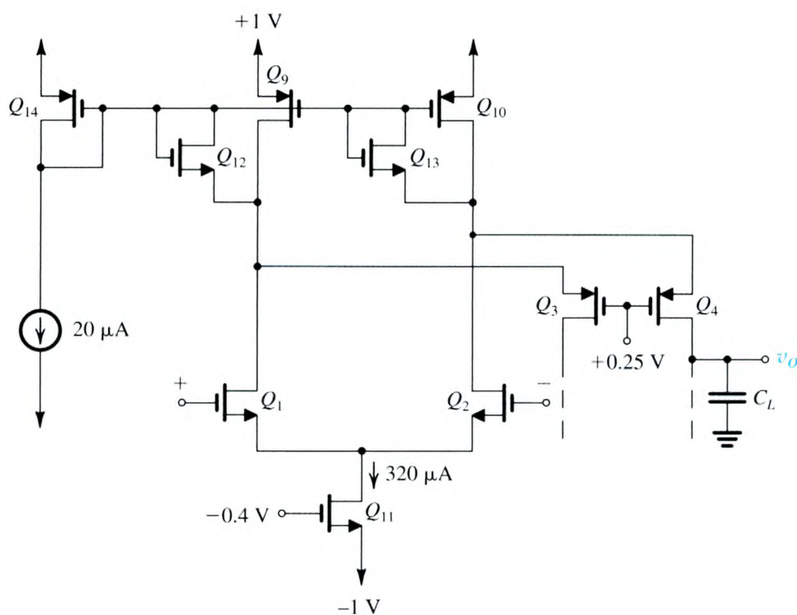


Figure P13.35

20 MHz when $C_L = 10$ pF. Design for $I_B = I$, and operate all devices at the same $|V_{OV}|$. Utilize transistors for which $|V_A|$ is specified to be 12 V. Find the required overdrive voltages and bias currents. What slew rate is achieved? Also, for the 0.25-μm CMOS technology of Appendix K in which $k_n' = 267\ \mu A/V^2$ and $k_p' = 93\ \mu A/V^2$, specify the required W/L of each of the 11 transistors used.

D 13.34 Sketch the circuit that is complementary to that in Fig. 13.10, that is, one that uses an input p-channel differential pair.

****13.35** This problem presents a very interesting addition to the folded-cascode op-amp circuit of Fig. 13.10, designed to deal with the situation during amplifier slewing. In particular, the additional circuitry does two things: It prevents Q_1 and Q_{11} from going into the triode region, and it increases the current available to charge C_L and thus increases the slew rate. The circuit is shown in Fig. P13.35 (with the current-mirror circuit omitted, for simplicity). Observe that three transistors are added: Q_{14}, which is biased by a constant-current source (20 μA), establishes the dc currents in Q_9 and Q_{10}. Assume with respect to Q_9 and Q_{10} that each

has a W/L ratio 10 times that of Q_{14}. The other two additional transistors are Q_{12} and Q_{13}, which are diode connected and are normally cut off.

(a) For $V_{id} = 0$, find the bias current in each of Q_1, Q_2, Q_3, Q_4, Q_{14}, Q_9, and Q_{10}. Also, for the dc voltages shown, and assuming $V_{tn} = |V_{tp}| = 0.45$ V and that all conducting devices are operating at $|V_{OV}| = 0.15$ V, show that Q_{12} and Q_{13} will be cut off.

(b) For an input differential signal that causes Q_2 to turn off and Q_1 to conduct the entire bias current (320 μA), Q_{12} turns on (while Q_{13} remains off). Noting that the drain current of Q_{12} adds to the 20 μA flowing through Q_{14}, find the current that now flows through Q_{10} and onto C_L. By what factor does the slew rate increase relative to the value without Q_{12} present? Also give an approximate estimate of the voltage at the drain of Q_1 during the slewing transient.

13.36 For the circuit in Fig. 13.12, assume that all transistors are operating at equal overdrive voltages of 0.15-V magnitude and have $|V_t| = 0.45$ V and that $V_{DD} = V_{SS} = 1$ V. Find (a) the range over which the NMOS input stage operates, (b) the range over which the PMOS input stage operates, (c) the range over which both operate (the overlap range), and (d) the input common-mode range. Assume that all current sources require a minimum voltage of $|V_{OV}|$ to operate properly.

13.37 A particular design of the wide-swing current mirror of Fig. 13.13(b) utilizes devices in the 65-nm CMOS technology of Appendix K having $W/L = 12$, $k_n' = 540\ \mu A/V^2$, and $V_t = 0.35$ V. For $I_{REF} = 90\ \mu A$, what value of V_{BIAS} is needed? Also give the voltages that you expect to appear at all nodes and specify the minimum voltage allowable at the output terminal. If V_A is specified to be 1 V, what is the output resistance of the mirror?

D 13.38 For the folded-cascode circuit of Fig. 13.9, let the total capacitance to ground at each of the source nodes of Q_3 and Q_4 be denoted C_P. Assuming that the incremental resistance between the drain of Q_3 and ground is small, show that the pole that arises at the interface between the first and second stages has a frequency $f_P \simeq g_{m3}/2\pi C_P$. Now, if this is the only nondominant pole, what is the largest value that C_P can be (expressed as a fraction of C_L) while a phase margin of 80° is achieved? Assume that all transistors are operated at the same bias current and overdrive voltage.

Section 13.3: BJT Op-Amp Techniques

Unless otherwise specified, for the problems in this section assume $\beta_N = 40$, $\beta_P = 10$, $V_{An} = 30$ V, $|V_{Ap}| = 20$ V, $|V_{BE}| = 0.7$ V, and $|V_{CEsat}| = 0.1$ V.

D 13.39 Design a Widlar current source shown in Fig. 13.14 so that all collector currents are 10 μA. Assume that $V_{CC} = 3$ V, $I_{S2} = 10^{-16}$, all other transistors have $I_S = 2 \cdot 10^{-17}$, and all have high β. Find V_{BIAS1}, V_{BIAS2}, and the values of all resistors. The voltage drops across R_3 and R_4 should be 0.1 V.

D 13.40 Design the circuit in Fig. 13.14 to generate a current $I = 5$ μA. Utilize transistors Q_1 and Q_2 having areas in a ratio of 1:4. Assume that Q_3 and Q_4 are matched and design for a 0.2-V drop across each of R_3 and R_4. Specify the values of R_2, R_3, and R_4. Ignore base currents.

D 13.41 Consider the circuit of Fig. 13.14 for the case designed in Exercise 13.14, namely, $I = 10$ μA, $I_{S2}/I_{S1} = 2$, $R_2 = 1.73$ kΩ, $R_3 = R_4 = 20$ kΩ. We must generate constant currents of 10 μA and 40 μA in Q_5 and Q_6, respectively. What should the emitter areas of Q_5 and Q_6 be relative to that of Q_1 with $R_5 = R_6 = 0$? What value of a resistance R_6 will, when connected in the emitter of Q_6, reduce the current generated by Q_6 to 10 μA? Assuming that the V_{BIAS1} line has a low incremental resistance to ground, find the output resistance of current source Q_5 and of current source Q_6 with R_6 connected. Ignore base currents.

D 13.42 It is required to use the current I_5 in Fig. 13.14 to bias an *npn* differential pair. The transistor Q_5 is identical to Q_2, and R_5 is equal to R_2. The differential pair has two equal collector resistances R_C connected to V_{CC}, and the output voltage v_O is taken between the two collectors.

(a) Find an expression for the differential gain A_d in terms of (R_C/R_5) and (I_{S5}/I_{S1}). Comment on the expected temperature dependence of A_d. Neglect the effect of finite β_N.

(b) Specify the values of R_5 and R_C required for $I = 20$ μA and $A_d = 6$ V/V. Let the emitter areas of Q_1 and Q_5 be in the ratio 1:4.

13.43 Consider the current-mirror active load $Q_{3,4}$ in the two-stage op amp of Fig. 13.15 with $R_1 = 400$ Ω. What value must R_1 be increased to in order to increase its output resistance by a factor of 2? Assume that $I_1 = 50\ \mu A$ and all transistors have $\beta = 200$ and $V_A = 125$ V.

***13.44** The resistors R_1 in Fig. 13.15 are mismatched so that one has a value $R_1 + \Delta R_1$. You may assume that the base currents are negligible.

(a) Show that this results in mismatched emitter currents in Q_1 and Q_2,

$$\frac{\Delta I_E}{I_E} = \frac{\Delta R_1}{R_1 + \Delta R_1 + r_e}$$

(b) Show that this, in turn, results in an input offset voltage of

$$V_{OS} = \frac{\Delta R_1 V_T}{R_1 + \Delta R_1 + r_e}$$

(c) If an input offset voltage of V_{OS} is observed, show that this can be compensated for (i.e., reduced to zero) by creating a relative mismatch in the resistors R_1,

$$\frac{\Delta R_1}{R_1} = \frac{V_{OS}}{V_T} \frac{1 + r_e/R_1}{1 - V_{OS}/V_T}$$

D 13.45 Design a Widlar current reference to generate the bias currents I_1 and I_2 of the op amp in Fig. 13.15. Use the circuit of Fig. 13.14 (excluding transistors Q_5 and Q_6) with $I = 15\ \mu A$, $I_7 = 45\ \mu A$, and $I_8 = 150\ \mu A$. Assume $I_{S1,3,4} = 10^{-15}\ \mu A$ and $I_{S2} = 4 \times 10^{-15}$. Specify the areas of Q_7 and Q_8 in terms of the area of Q_3 and all resistor values. Design for a voltage drop of 0.1 V across R_3 and R_4.

13.46 For the op-amp circuit in Fig. 13.15, assume that $V_{CC} = 5$ V, $|V_{BE}| = 0.6$ V for all transistors, and $R_2 = 300\ \Omega$. Also assume that the current source $I_2 = 200\ \mu A$ requires a voltage drop of at least 0.3 V to operate properly. Find the resulting limits on the output swing.

D 13.47 (a) Find the input common-mode range of the circuit in Fig. 13.15 assuming $R_1 = 0$ and the current source I_1 is simply a *pnp* transistor biased in active mode with $V_{CC} = 3$ V.
(b) Give the complementary version of the input stage in Fig. 13.15; that is, the one in which the differential pair is *npn*. For the same conditions as in (a), what is the input common-mode range?

13.48 For the circuit in Fig. 13.16, let $V_{CC} = 3$ V, $V_{BIAS} = 2.3$ V, $I = 20\ \mu A$, and $R_C = 35$ kΩ. Find the input common-mode range and the differential voltage gain v_o/v_{id}. Neglect base currents.

D 13.49 For the circuit in Fig. 13.17, let $V_{CC} = 3$ V, $V_{BIAS} = 0.7$ V, and $I_{C6} = 40\ \mu A$. Find R_C that results in a differential gain of 14 V/V. What is the input common-mode range and the input differential resistance? Ignore base currents except when calculating R_{id}. If R_{id} is to be increased by a factor of 4 while the gain and V_{ICM} remain unchanged, what must I and R_C be changed to?

13.50 It is required to find the input resistance and the voltage gain of the input stage shown in Fig. 13.18. Let $V_{ICM} \ll 0.8$ V so that the Q_3–Q_4 pair is off. Assume that Q_5 supplies 8 μA, that each of Q_7 to Q_{10} is biased at 8 μA, and that all four cascode transistors are operating in the active mode. The input resistance of the second stage of the op amp is 1.5 MΩ. The emitter-degeneration resistances are $R_7 = R_8 = 22$ kΩ, and $R_9 = R_{10} = 33$ kΩ. [*Hint*: Refer to Fig. 13.19.]

D *13.51 Consider the equivalent half-circuit shown in Fig. 13.19. Assume that in the original circuit, Q_1 is biased at a current I, Q_7 and Q_9 are biased at $2I$, the dc voltage drop across R_7 is 0.2 V, and the dc voltage drop across R_9 is 0.3 V. Find the output resistance in terms of I, and hence find the open-circuit voltage gain (i.e., the voltage gain for $R_L = \infty$). Now with R_L connected, find the voltage gain in terms of (IR_L). For $R_L = 1$ MΩ, find I that will result in the voltage gains of 150 V/V and 300 V/V.

***13.52** (a) For the circuit in Fig. 13.20, show that the loop gain of the common-mode feedback loop is

$$A\beta \simeq \frac{R_{o9} \| R_{o7}}{r_{e7} + R_7}$$

Recall that the CMF circuit responds only to the average voltage V_{CM} of its two input voltages and realizes the transfer characteristic $V_B = V_{CM} + 0.4$. Ignore the loading effect of the CMF circuit on the collectors of the cascode transistors.
(b) For the values in Example 13.4, calculate the loop gain $A\beta$.
(c) In Example 13.4, we found that with the CMF absent, a current mismatch $\Delta I = 0.3\ \mu A$ gives rise to $\Delta V_{CM} = 2.5$ V. Now, with the CMF present, use the value of loop gain found in (b) to calculate the expected ΔV_{CM} and compare to the value found by a different approach in Example 13.4. [*Hint*: Recall that negative feedback reduces change by a factor equal to $(1 + A\beta)$.]

13.53 In the circuit of Fig. 13.22, Q_1 and Q_2 exhibit emitter–base breakdown at 7 V, while for Q_3 and Q_4 such a breakdown occurs at about 50 V. What differential input voltage would result in the breakdown of the input-stage transistors?

13.54 Consider the dc analysis of the 741 input stage shown in Fig. 13.23.

(a) Derive an expression for I taking β_P into account. What is the percentage change in I if β_P drops from 50 to 20?
(b) Now, consider an alternative design of this circuit in which the feedback loop is eliminated. That is, Q_5 and Q_6 are eliminated, and I_{REF} is fed to the common-base connection of Q_3 and Q_4. What is I now in terms of I_{REF}? If β_P changes from 50 to 20, what is the resulting percentage change in I?

D 13.55 Consider the dc analysis of the 741 input stage shown in Fig. 13.23 for the situation in which $I_{S6} = 2I_{S5}$. For $I_{REF} = 19\ \mu A$ and assuming β_P to be high, what does I become? Design a Widlar source to reestablish $I_{C1} = I_{C2} = 9.5\ \mu A$ using the circuit in Fig. 13.14 wherein all resistors and transistors are sized identically except Q_1 in Fig. 13.14, whose emitter area is sized one-tenth that of the others.

D 13.56 Consider the circuit shown in Fig. 13.23. If $I_{REF} = 40\ \mu A$ and I is required to be $10\ \mu A$, what must be the ratio of the emitter–junction area of Q_6 to that of Q_5? Assume that β_P is large.

13.57 Consider the input circuit of the 741 op amp of Fig. 13.23 when the emitter current of Q_5 is about $40\ \mu A$. If β of Q_1 is 150 and that of Q_2 is 220, find the input bias current I_B and the input offset current I_{OS} of the op amp.

13.58 The circuit of Fig. 13.23 has an emitter current for Q_5 of $40\ \mu A$ with input bias and offset currents limited to 100 nA and 10 nA, respectively. What minimum β_N and what β_N variation are implied?

13.59 Consider the 741 input stage as modeled in Fig. 13.24, with two additional *npn* diode-connected transistors, Q_{1a} and Q_{2a}, connected between the present *npn* and *pnp* devices, one per side. Assume $\beta_N = 200$, $\beta_P = 50$, and $V_{Ap} = 50$ V. Convince yourself that each of the additional devices will be biased at the same current as Q_1 to Q_4—let us assume $9.5\ \mu A$. What does R_{id} become? What does G_{m1} become? What is the value of R_{o4} now? What is the output resistance of the first stage, R_{o1}? (Assume the active load output resistance remains unchanged from Exercise 13.21 at 10.5 MΩ.) What is the new open-circuit voltage gain, $G_{m1}R_{o1}$? Compare these values with the original ones from Exercises 13.20–22, namely, $R_{id} = 2.1$ MΩ, $G_{m1} = 0.19$ mA/V, $R_{o4} = 10.5$ MΩ, $R_{o1} = 5.3$ MΩ, and $|A_{vo}| = 1007$ V/V.

13.60 If the current transfer ratio of the mirror load of the 741 input stage is 0.995, find the CMRR of the input stage. The input transistors are biased at $9.5\ \mu A$. (*Hint:* Use Eq. 13.86 together with the output resistance values determined in Exercise 13.2.)

13.61 Consider the circuit of Fig. 13.27 modified to include resistors R in series with the emitters of each of Q_5 and Q_6. Assuming $|V_A| = 50$ V, $\beta_p = 50$, and $I_{C6} = 20\ \mu A$, what value of R provides $R_{o6} = 20$ MΩ? How does this affect the CMRR calculated in Exercise 13.24?

13.62 An alternative approach to that presented in Example 13.5 for determining the CMRR of the 741 input stage is investigated in this problem. Rather than performing the analysis on the closed loop shown in Fig. 13.27, we observe that the negative feedback increases the resistance at node Y by the amount of negative feedback. Thus, we can break the loop at Y and connect a resistance $R_f = (1 + A\beta)R_o$ between the common-base connection of Q_3–Q_4 and ground. We can then determine the current i and G_{mcm}. Using the fact that the loop gain is approximately equal to β_P (Exercise 13.17) show that this approach yields an identical result to that found in Example 13.5.

D 13.63 For the circuit in Fig. P13.63, neglect base currents and use the exponential i_C–v_{BE} relationship to show that

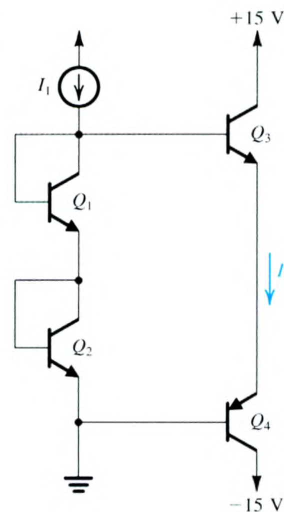


Figure P13.63

$$I_3 = I_1\sqrt{\frac{I_{S3}I_{S4}}{I_{S1}I_{S2}}}$$

Find I_1 for the case in which $I_{S3} = I_{S4} = 3 \times 10^{-14}$ A, $I_{S1} = I_{S2} = 10^{-14}$ A, and a bias current $I_3 = 150\ \mu A$ is required.

D 13.64 Figure P13.64 shows the CMOS version of the circuit in Fig. P13.63. Find the relationship between I_3 and I_1

in terms of k_1, k_2, k_3, and k_4 of the four transistors, assuming the threshold voltages of all devices to be equal in magnitude. Note that k denotes $\mu C_{ox}W/L$. In the event that $k_1 = k_2$ and $k_3 = k_4 = 20k_1$, find the required value of I_1 to yield a bias current in Q_3 and Q_4 of 0.2 mA.

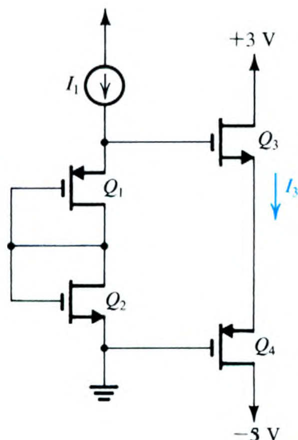


Figure P13.64

D *13.65 The V_{BE}– multiplier circuit shown in Fig. P13.65 can replace the two series diode-connected transistors in Fig. 13.15 to establish the voltage drop needed to bias the output stage transistors. Design the circuit to provide a terminal voltage of 1.2 V. Base your design on half the current flowing through R_1, and assume that $I_S = 10^{-14}$ A and $\beta = 200$. What is the incremental resistance between the two terminals of the V_{BE}– multiplier circuit?

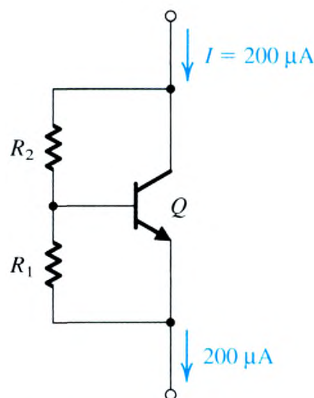


Figure P13.65

SIM D 13.66 Figure P13.66 shows a circuit suitable for op-amp applications. For all transistors $\beta = 100$, $V_{BE} = 0.7$ V, and $r_o = \infty$.

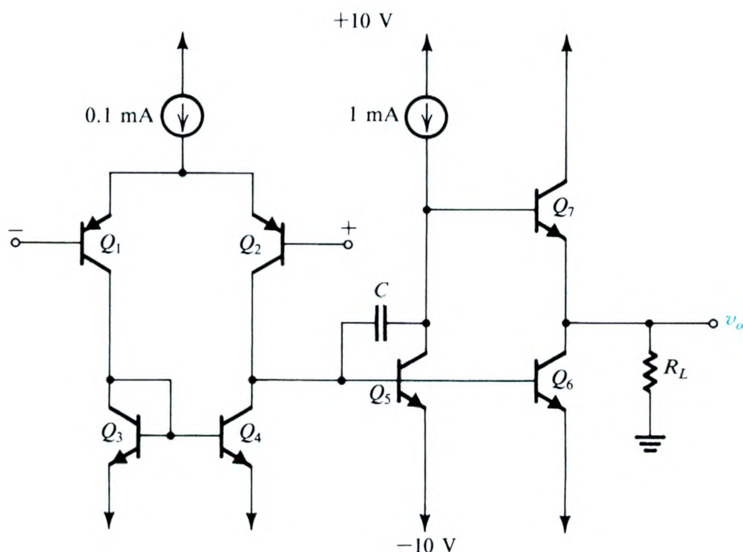


Figure P13.66

(a) For inputs grounded and output held at 0 V (by negative feedback) find the collector currents of all transistors. Neglect base currents.
(b) Calculate the input resistance.
(c) Calculate the gain of the amplifier with a load of 5 kΩ.
(d) With load as in (c) calculate the value of the capacitor C required for a 3-dB frequency of 100 Hz.

13.67 The output stage in Fig. 13.28 operates at a quiescent current I_Q of 0.6 mA. The maximum current i_L that the stage can provide in either direction is 12 mA. The output stage is equipped with a feedback circuit that maintains a minimum current of $I_Q/2$ in the inactive output transistor. Also, $V_{CC} =$ 3 V.

(a) What is the allowable range of v_O?
(b) For $i_L = 0$, what is the output resistance of the op amp?
(c) If the open-loop gain of the op amp is 100,000 V/V, find the closed-loop output resistance obtained when the op amp is connected in the unity-gain voltage follower configuration, with $i_L = 0$.
(d) If the op amp is sourcing a load current $i_L = 12$ mA, find i_P, i_N, and the open-loop output resistance.
(e) Repeat (d) for the case of the open-loop op amp sinking a load current of 12 mA.

13.68 It is required to derive the expressions in Eqs. (13.87) and (13.88). Toward that end, first find v_{B7} in terms of v_{BEN} and hence i_N. Then find v_{B6} in terms of i_P. For the latter purpose note that Q_4 measures v_{EBP} and develops a current $i_4 = (v_{EBP} - v_{EB4})/R_4$. This current is supplied to the series connection of Q_5 and R_5 where $R_5 = R_4$. In the expression you obtain for v_{B6}, use the relationship

$$\frac{I_{SP}}{I_{S4}} = \frac{I_{SN}}{I_{S5}}$$

to express v_{B6} in terms of i_P and I_{SN}. Now with v_{B6} and v_{B7} determined, find i_{C6} and i_{C7}.

13.69 It is required to derive the expression for v_E in Eq. (13.89). Toward that end, note from the circuit in Fig. 13.30 that $v_E = v_{EB7} + v_{BEN}$ and note that Q_N conducts a current i_N and Q_7 conducts a current i_{C7} given by Eq. (13.88).

D 13.70 For the output stage in Fig. 13.30, find the current I_{REF} that results in a quiescent current $I_Q = 0.6$ mA. Assume that $I = 12$ μA, Q_N has eight times the area of Q_{10}, and Q_7 has four times the area of Q_{11}. What is the minimum current in Q_N and Q_P?

CHAPTER 14

Filters

IN THIS CHAPTER YOU WILL LEARN

- How filters are characterized by their signal-transmission properties and classified into different types based on the relative location of their passband(s) and stopband(s).
- How filters are specified and how to obtain a filter transfer function that meets the given specifications, including the use of popular special functions such as the Butterworth and the Chebyshev.
- The various first-order and second-order filter functions and their realization using op amps and RC circuits.
- The basic second-order LCR resonator and how it can be used to realize the various second-order filter functions.
- The best op amp–RC circuit for realizing an inductance and how it can be used to realize the various second-order filter functions.
- That connecting two op-amp integrators, one inverting and one noninverting, in a feedback loop realizes a second-order resonance circuit and can be used to obtain circuit realizations of the various second-order filter functions.
- How second-order filter functions can be realized using a single op amp and an RC circuit.
- The basis for the most popular approach to the realization of filter functions in IC form: the switched-capacitor technique.

Introduction

In this chapter, we study the design of an important building block of communications and instrumentation systems, the electronic filter. Filter design is one of the very few areas of engineering for which a complete design theory exists, starting from specification and ending with a circuit implementation. A detailed study of filter design requires an entire book, and indeed such textbooks exist. In the limited space available here, we will concentrate on a selection of topics that provide an introduction to this rich subject. The material presented here together with the design material and tables included in Appendix H (available on the website) provide a significant resource for filter circuits and design methods.

The filters studied in this chapter are *analog*, and they find extensive application despite the widespread use of digital signal processing. This is because most systems have to interface with the physical world, which is analog in nature. Such an interface, at a minimum, consists of an amplifier and an analog filter.

The material in this chapter makes extensive use of poles and zeros (Appendix F), and requires a thorough familiarity with op-amp circuit applications (Chapter 2).

14.1 Basic Filter Concepts

14.1.1 Filter Transmission

The filters we are about to study are linear circuits that can be represented by the general two-port network shown in Fig. 14.1. The filter **transfer function** $T(s)$ is the ratio of the output voltage $V_o(s)$ to the input voltage $V_i(s)$,

$$T(s) \equiv \frac{V_o(s)}{V_i(s)} \tag{14.1}$$

The filter **transmission** is found by evaluating $T(s)$ for *physical frequencies*, $s = j\omega$, and can be expressed in terms of its magnitude and phase as

$$T(j\omega) = |T(j\omega)|e^{j\phi(\omega)} \tag{14.2}$$

The magnitude of transmission is often expressed in decibels in terms of the **gain function**

$$G(\omega) \equiv 20\log|T(j\omega)|, \text{dB} \tag{14.3}$$

or, alternatively, in terms of the **attenuation function**

$$A(\omega) \equiv -20\log|T(j\omega)|, \text{dB} \tag{14.4}$$

A filter shapes the frequency spectrum of the input signal, $|V_i(j\omega)|$, according to the magnitude of the transfer function $|T(j\omega)|$, thus providing an output $V_o(j\omega)$ with a spectrum

$$|V_o(j\omega)| = |T(j\omega)||V_i(j\omega)| \tag{14.5}$$

Also, the phase characteristics of the signal are modified as it passes through the filter according to the filter phase function $\phi(\omega)$.

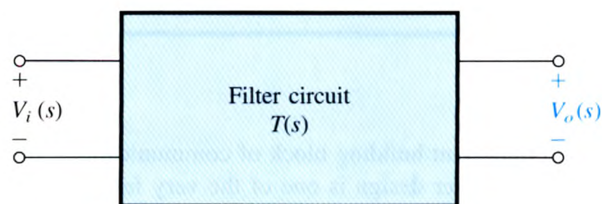


Figure 14.1 The filters studied in this chapter are linear circuits represented by the general two-port network shown. The filter transfer function $T(s) \equiv V_o(s)/V_i(s)$.

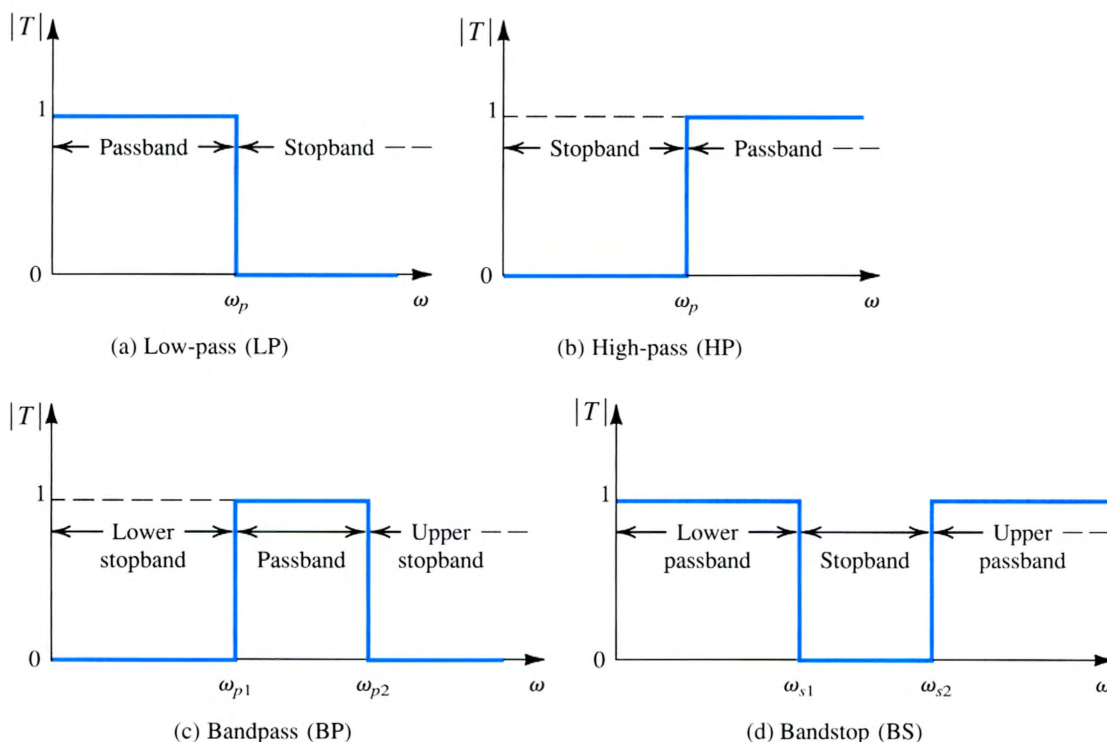


Figure 14.2 Ideal transmission characteristics of the four major filter types: **(a)** low-pass (LP), **(b)** high-pass (HP), **(c)** bandpass (BP), and **(d)** bandstop (BS).

14.1.2 Filter Types

We are specifically interested here in filters that perform a **frequency-selection** function: **passing** signals whose frequency spectrum lies within a specified range, and **stopping** signals whose frequency spectrum falls outside this range. Such a filter has ideally a frequency band (or bands) over which the magnitude of transmission is unity (the filter **passband**) and a frequency band (or bands) over which the transmission is zero (the filter **stopband**). Figure 14.2 depicts the ideal transmission characteristics of the four major filter types: **low-pass** (LP) in Fig. 14.2(a), **high-pass** (HP) in Fig. 14.2(b), **bandpass** (BP) in Fig. 14.2(c),

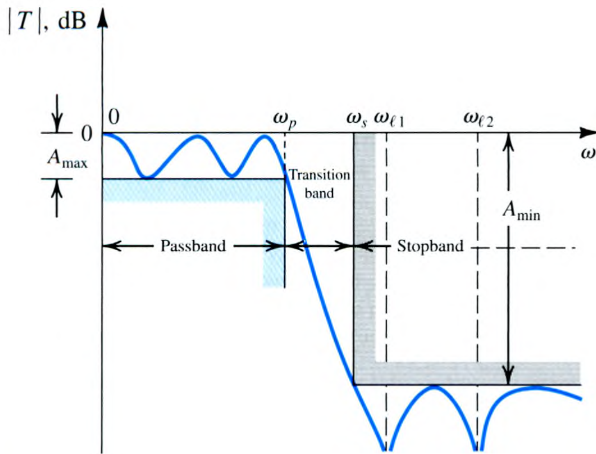


Figure 14.3 Specification of the transmission characteristics of a low-pass filter. The magnitude response of a filter that just meets specifications is also shown.

and **bandstop** (BS) or **band-reject** in Fig. 14.2(d). These idealized characteristics, by virtue of their vertical edges, are known as **brick-wall** responses.

14.1.3 Filter Specification

The filter-design process begins with the filter user specifying the transmission characteristics required of the filter. Such a specification cannot take the form shown in Fig. 14.2 because physical circuits cannot realize these idealized characteristics. Figure 14.3 shows realistic specifications for the transmission characteristics of a low-pass filter. Observe that since a physical circuit cannot provide constant transmission at all passband frequencies, the specifications allow for deviation of the passband transmission from the ideal unity or 0 dB, but place an upper bound, A_{max} (dB), on this deviation. Depending on the application, A_{max} typically ranges from 0.05 dB to 3 dB. Also, since a physical circuit cannot provide zero transmission (infinite attenuation) at all stopband frequencies, the specifications in Fig. 14.3 allow for some transmission over the stopband. However, the specifications require the stopband signals to be attenuated by at least A_{min} (dB) relative to the passband signals. Depending on the filter application, A_{min} can range from 20 dB to 100 dB.

Since the transmission of a physical circuit cannot change abruptly at the edge of the passband, the specifications of Fig. 14.3 provide for a band of frequencies over which the attenuation increases from near 0 dB to A_{min}. This **transition band** extends from the passband edge ω_p to the stopband edge ω_s. The ratio ω_s/ω_p is usually used as a measure of the sharpness of the low-pass filter response and is called the **selectivity factor**. Finally, observe that for convenience the passband transmission is specified to be 0 dB. The final filter, however, can be given a passband gain, if desired, without changing its selectivity characteristics.

To summarize, the transmission of a low-pass filter is specified by four parameters:

1. The passband edge ω_p
2. The maximum allowed variation in passband transmission A_{max}
3. The stopband edge ω_s
4. The minimum required stopband attenuation A_{min}

ANALOG FILTERS: A BRIEF HISTORY

Driven by the needs of the emerging telephone system in the early twentieth century, the earliest analog filters were designed using ad hoc methods with inductors (coils) and capacitors. It was not until the 1940s that a design theory for passive LC filters was developed. This filter synthesis method, however, required extensive computation and had to await the emergence of sufficiently advanced computer technology in the 1950s and the 1960s to become adopted, further developed, and utilized.

The invention of the transistor and the IC resulted in new ways of assembling electronic circuits. However, these were incompatible with the bulky and heavy inductors used in audio-frequency LC filters. The search for inductorless filters thus began and was greatly influenced by the emergence of the low-cost IC op amp in the late 1960s. Extensive research in the 1960s and 1970s resulted in a large repertoire of op amp–RC filters, the best of which are studied in this chapter.

No sooner had op amp–RC filters become a mature and reliable technology than the need arose for analog filters that could be fully integrated. Multiple approaches were proposed in the late 1970s, but the switched-capacitor circuit has become the technology of choice for low-frequency applications, mostly because of its compatibility with CMOS. For higher-frequency applications approaching the gigahertz range, needed for the burgeoning mobile communications market, two approaches have gained momentum: transconductance-C and active-LC. The latter uses actual physical inductors (in the nanohenry range) fabricated on the IC chip.

The more tightly one specifies a filter—that is, lower A_{max}, higher A_{min}, and/or a selectivity ratio ω_s/ω_p closer to unity—the closer the response of the resulting filter will be to the ideal. However, the resulting filter transfer function will be of higher order, and thus the filter circuit will be more complex and expensive.

In addition to the magnitude of transmission, the phase response of the filter is of interest in some applications. The filter-design problem, however, is considerably complicated when both magnitude and phase are to be specified.

14.1.4 Obtaining the Filter Transfer Function: Filter Approximation

Once the filter specifications have been decided upon, the next step in the design is to find a transfer function whose magnitude meets the specification. To meet specification, the magnitude-response curve must lie in the unshaded area in Fig. 14.3. The curve shown in the figure is for a filter that *just* meets specifications. Observe that for this particular filter, the magnitude response *ripples* throughout the passband, and the ripple peaks are all equal. Since the peak ripple is equal to A_{max}, it is usual to refer to A_{max} as the **passband ripple**. The particular filter response shows ripples also in the stopband, again with the ripple peaks all equal and of such a value that the minimum stopband attenuation achieved is equal to the specified value, A_{min}. Thus this particular response is said to be **equiripple** in both the passband and the stopband.

The process of obtaining a transfer function that meets given specifications is known as **filter approximation**. Filter approximation is usually performed using computer programs (Snelgrove, 1982; Ouslis and Sedra, 1995) or filter-design tables (Zverev, 1967). In simpler cases, filter approximation can be performed using closed-form expressions, as will be seen in Section 14.3.

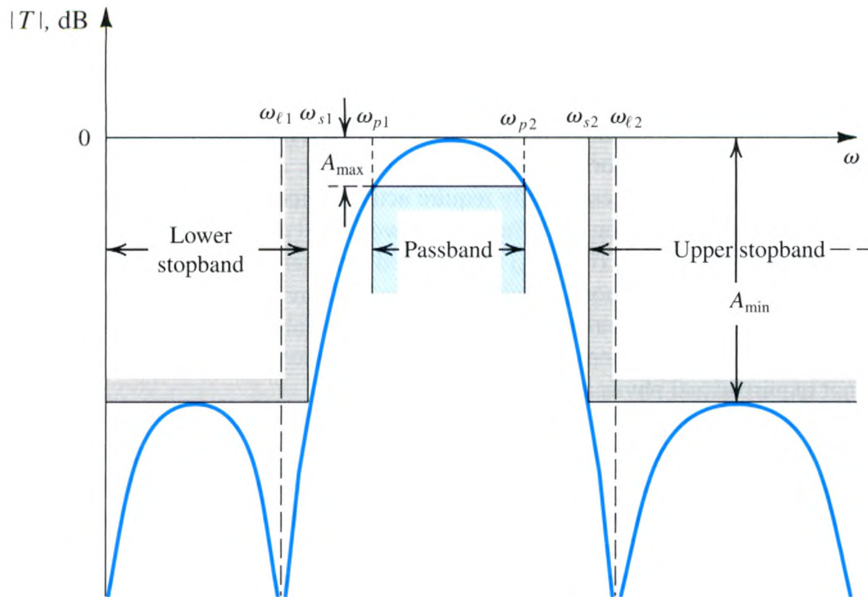


Figure 14.4 Transmission specifications for a bandpass filter. The magnitude response of a filter that just meets specifications is also shown. Note that this particular filter has a monotonically decreasing transmission in the passband on both sides of the peak frequency.

As another example, Fig. 14.4 shows transmission specifications for a bandpass filter and the response of a filter that meets these specifications. For this example we have chosen an approximation function that does not ripple in the passband; rather, the transmission decreases monotonically on both sides of a center frequency, attaining the maximum allowable deviation at the two edges of the passband.

Filter transfer functions have special characteristics. We will study these characteristics in Section 14.2.

EXERCISES

14.1 Find approximate values of attenuation (in dB) corresponding to filter transmissions of 1, 0.99, 0.9, 0.8, 0.7, 0.5, 0.1, 0.
Ans. 0, 0.1, 1, 2, 3, 6, 20, ∞ (dB)

14.2 If the variation in the magnitude of passband transmission is to be limited to 10%, and if the stopband transmission is to be no greater than 1% of the passband transmission, find A_{max} and A_{min}.
Ans. 0.9 dB; 40 dB

14.1.5 Obtaining the Filter Circuit: Filter Realization

Having obtained a suitable transfer function $T(s)$, the next step in the filter design process is to find a circuit whose transfer function is equal to $T(s)$. This is known as **filter realization**.

There are a number of technologies and methods for the realization of analog filters. The oldest technology uses inductors, capacitors, and resistors, and the resulting LCR circuits are **passive**, since they do not essentially require active elements (transistors or op amps) to realize the filter transfer function $T(s)$. Such filters work well at high frequencies; however, in the low-frequency band (dc to 100 kHz) the required inductances are large and the inductors are physically bulky. Furthermore, inductors of this kind are incompatible with the modern techniques for assembling electronic systems and are certainly impossible to manufacture in monolithic form. Therefore, there has been considerable interest in finding filter realizations that do not require actual physical inductors.

A popular approach for the design of **inductorless filters** consists of replacing each inductance in the LCR filter with a circuit composed of op amps, resistors, and a capacitor, and having an input impedance equal to sL. Such a circuit is said to **simulate** the inductance L. We will study passive LCR filters in Section 14.4 and the design of op amp–RC filters based on inductance simulation in Section 14.5.

The op amp–RC filter circuits in Section 14.5 represent a special class of **active filters**. In Section 14.6 we will study another important type of active filter, based on the use of op-amp integrators. And in Section 14.7 we will consider a third kind of active filter that uses a single op amp together with resistors and capacitors to realize a second-order filter function.

The active filter circuits presented in Sections 14.5–14.7 can be built using discrete op amps, resistors, and capacitors assembled on printed-circuit boards. Large-volume production employs special technologies using thin-film and thick-film components. However, none of these technologies yields the economics achieved by monolithic integrated-circuit fabrication.

Today there are two popular approaches for realizing fully integrated analog filters: the **switched-capacitor** approach, which we study in Section 14.8, and the **transconductance** $-C$ ($G_m - C$) approach presented on the website. The latter is particularly well suited for high-frequency applications.

Switched-capacitor filters use CMOS op amps together with MOS switches and capacitors. As we will see in Section 14.8, the signals are processed at discrete-time intervals, and so the resulting circuits are **discrete-time** circuits. However, because the signal magnitude is continuous (not quantized), these are still analog filters.

14.2 The Filter Transfer Function

The filter transfer function $T(s)$ can be written as the ratio of two polynomials:

$$T(s) = \frac{a_M s^M + a_{M-1} s^{M-1} + \cdots + a_0}{s^N + b_{N-1} s^{N-1} + \cdots + b_0} \tag{14.6}$$

where the numerator and denominator coefficients are real numbers. The numerator and denominator polynomials can be factored out, and $T(s)$ expressed in terms of its poles and zeroes:

$$T(s) = a_M \frac{(s - z_1)(s - z_2) \cdots (s - z_M)}{(s - p_1)(s - p_2) \cdots (s - p_N)} \tag{14.7}$$

Next, we will consider some important characteristics of the transfer functions of the filters studied in this chapter.

14.2.1 The Filter Order

The filter order is equal to the degree N of the denominator polynomial and thus is equal to the number of poles. We will see, shortly, that the filter order N is also equal to the number of transmission zeros, provided we include in our count the transmission zeros at $s = \infty$ as well as, of course, the finite transmission zeros z_1 to z_M.

14.2.2 The Filter Poles

For a filter, or indeed any circuit or system, to be stable, all the poles must lie in the left half of the s plane. Also, since the coefficients of the denominator polynomial (i.e., b_0, b_1, ..., b_{N-1}) are real numbers, the poles must either lie on the negative real axis or occur in complex-conjugate pairs.

To obtain highly selective (i.e., sharp) filter responses, the poles are usually complex conjugate except for one real pole if the filter order N is odd. In fact, the sharper the response a filter is required to have, the closer its complex poles must be to the vertical $j\omega$-axis of the s plane. For example, Fig. 14.5 shows the pole locations for the fifth-order ($N = 5$) low-pass filter whose transmission function is depicted in Fig. 14.3. Note that this filter has two pairs of complex-conjugate poles and one negative real-axis pole, for a total of five poles. All the poles lie in the vicinity of the filter passband, which is what gives the filter its high transmission at passband frequencies.

As another example, consider the sixth-order ($N = 6$) bandpass filter whose transmission function $|T(j\omega)|$ is shown in Fig. 14.4. Its pole locations are shown in Fig. 14.6. Notice that there are three pairs of complex-conjugate poles. Here, also, the poles lie close to the $j\omega$-axis, in the vicinity of the filter passband.

14.2.3 The Filter Transmission Zeros

The filter transmission zeros determine the filter type (i.e., low-pass, high-pass, etc.). This is so because the transmission zeros are located at stopband frequencies including 0 and ∞. Transmission zeros at $\omega = 0$ give rise to factors of the form s^k in the numerator of the transfer function, where k is the number of transmission zeros at $s = 0$ ($\omega = 0$). High-pass and bandpass filters usually have at least one transmission zero at $s = 0$. Of course, low-pass filters cannot have zeros at $s = 0$ as they are required to pass signals with zero frequency (dc).

Transmission zeros at $s = \infty$ manifest themselves as the difference between the order M of the numerator polynomial and the order N of the denominator polynomial. This is the case because, as we can see from Eq. (14.6), as s approaches ∞, $T(s)$ approaches zero in the manner of a_M/s^{N-M}; the filter is thus said to have $(N - M)$ transmission zeros at $s = \infty$. Note that M cannot exceed N if the filter is to be stable; otherwise as s approaches ∞, $T(s)$ approaches ∞, which is not possible. A high-pass filter will have $M = N$, resulting in a finite transmission equal to a_M at high frequencies. Low-pass and bandpass filters, on the other hand, usually have at least one transmission zero at $s = \infty$, and so M will be less than N by at least one.

Apart from the transmission zeros at $s = 0$ and $s = \infty$, the filter may have transmission zeros at finite frequencies. For maximum effect, these finite zeros are placed at stopband

frequencies. Thus, a filter might have a transmission zero at a frequency ω_l, where ω_l is in the filter stopband. In this case, the numerator of the transfer function will have a factor $(s - j\omega_l)$. However, since complex zeros must occur in complex-conjugate pairs, the numerator must also have a factor $(s + j\omega_l)$. In other words, the numerator will have a factor $(s^2 + \omega_l^2)$, which for physical frequencies $s = j\omega$ becomes $(-\omega^2 + \omega_l^2)$, indicating zero transmission (infinite attenuation) at $\omega = \omega_l$, as desired.

As an example, consider once more the fifth-order low-pass filter whose transmission is shown in Fig. 14.3. The figure shows that the filter has infinite attenuation[1] and thus transmission zeros at $\omega = \omega_{l1}$, $\omega = \omega_{l2}$, and at $\omega = \infty$. That is, the transmission zeros are at $s = \pm j\omega_{l1}$, $s = \pm j\omega_{l2}$, and $s = \infty$, with all five zeros in the filter stopband. This means that the transfer function for this filter will be

$$T(s) = \frac{a_4\,(s^2 + \omega_{l1}^2)\,(s^2 + \omega_{l2}^2)}{s^5 + b_4\,s^4 + b_3\,s^3 + b_2\,s^2 + b_1\,s + b_0} \tag{14.8}$$

Figure 14.5 shows the pole and zero locations for this low-pass filter.

As another example, consider the sixth-order bandpass filter whose magnitude of transmission is shown in Fig. 14.4. This filter has transmission zeros at $s = \pm j\omega_{l1}$ and $s = \pm j\omega_{l2}$; one zero at $s = 0$; and one zero at $s = \infty$ (because $|T|$ tends to zero at both $\omega = 0$ and $\omega = \infty$). This makes for a total of six zeros (equal to N). Thus, the transfer function of this filter will be

$$T(s) = \frac{a_5\,s(s^2 + \omega_{l1}^2)\,(s^2 + \omega_{l2}^2)}{s^6 + b_5\,s^5 + b_4\,s^4 + b_3\,s^3 + b_2\,s^2 + b_1\,s + b_0} \tag{14.9}$$

Figure 14.6 shows the s-plane locations of the poles and zeros for this filter.

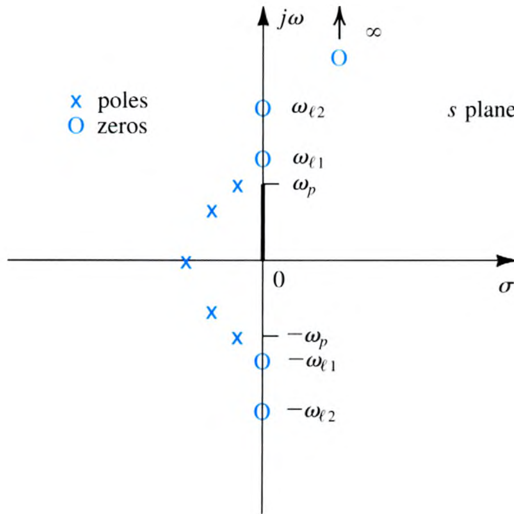


Figure 14.5 Pole–zero pattern for the low-pass filter whose transmission is sketched in Fig. 14.3. This is a fifth-order filter ($N = 5$).

[1] Note that when $|T| = 0$, $20 \log |T| = -\infty$.

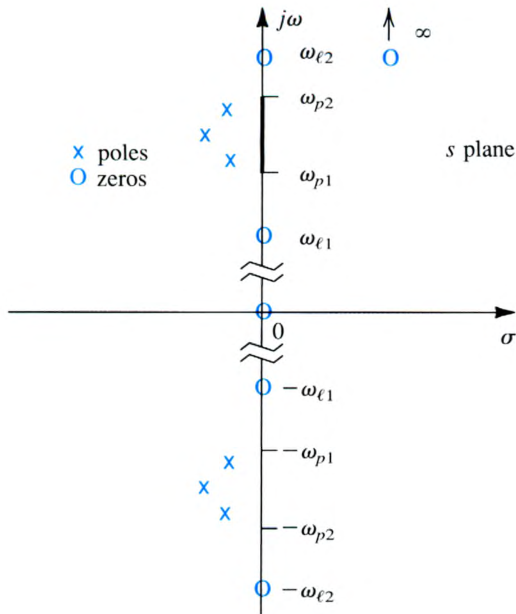


Figure 14.6 Pole–zero pattern for the bandpass filter whose transmission function is shown in Fig. 14.4. This is a sixth-order filter ($N = 6$).

Example 14.1

Give the order N, the type, and the transfer function $T(s)$ of a filter having the following poles and zeros:

$$\text{Poles}: s = -1, \;\; s = -0.5 \pm j0.7$$
$$\text{Zeros}: s = \infty, \;\; s = \pm j1.2$$

The transmission of the filter at dc is unity.

Solution

This is a low-pass filter with $N = 3$ and $T(s)$ of

$$T(s) = K \frac{(s + j1.2)\,(s - j1.2)}{(s + 1)\,(s + 0.5 - j0.7)\,(s + 0.5 + j0.7)}$$
$$= K \frac{s^2 + 1.44}{s^3 + 2\,s^2 + 1.74\,s + 0.74}$$

Since the dc ($s = 0$) transmission is unity, K can be found from

$$K \frac{1.44}{0.74} = 1 \Rightarrow \;\; K = 0.514$$

EXERCISE

14.3 Find the location of the transmission zeros and the type of the third-order filter whose numerator polynomial is:
(a) K, (b) Ks, (c) Ks^2, and (d) Ks^3, where K is a constant.
Ans. (a) three zeros at $s = \infty$, low-pass; (b) one zero at $s = 0$ and two zeros at $s = \infty$, bandpass; (c) two zeros at $s = 0$ and one zero at $s = \infty$, bandpass; (d) three zeros at $s = 0$, high-pass.

14.2.4 All-Pole Filters

Some low-pass filters have all their transmission zeros located at $s = \infty$. Thus, the magnitude of their transmission, $|T(j\omega)|$, decreases monotonically toward 0 as ω approaches ∞. Figure 14.7(a) shows the response of such a filter ($N = 5$ in this case). The transfer function for these filters takes the form

$$T(s) = \frac{a_0}{s^N + b_{N-1}s^{N-1} + \cdots + b_0} \tag{14.10}$$

Such a filter is known as an **all-pole filter**. Typical pole-zero locations for the fifth-order all-pole low-pass filter of Fig. 14.7(a) are shown in Fig. 14.7(b).

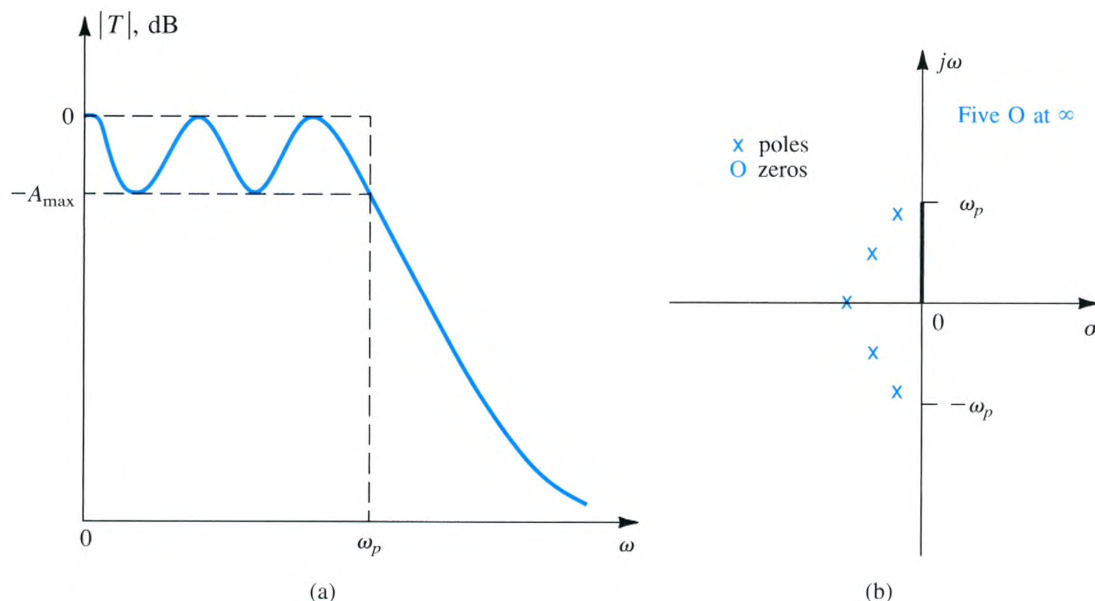


Figure 14.7 **(a)** Transmission magnitude of a fifth-order low-pass filter having all transmission zeros at infinity. **(b)** Pole–zero pattern for the filter in **(a)**.

EXERCISES

14.4 A second-order filter has its poles at $s = -(1/2) \pm j\left(\sqrt{3}/2\right)$. The transmission is zero at $\omega = 2$ rad/s and is unity at dc ($\omega = 0$). Find the transfer function.

Ans. $T(s) = \frac{1}{4}\frac{s^2+4}{s^2+s+1}$

14.5 A fourth-order filter has zero transmission at $\omega = 0$, $\omega = 2$ rad/s, and $\omega = \infty$. The poles are $-0.1 \pm j0.8$ and $-0.1 \pm j1.2$. Find $T(s)$.

Ans. $T(s) = \frac{a_3 s\left(s^2+4\right)}{(s^2+0.2s+0.65)(s^2+0.2s+1.45)}$

14.2.5 Factoring $T(s)$ into the Product of First-Order and Second-Order Functions

A simple, effective method for realizing high-order (i.e., orders greater than 2) active filters is to realize $T(s)$ by cascading second-order filter sections and, for odd N, a first-order filter section. Assuming that the output of each filter section is a low-impedance node, such as the output terminal of an op amp, the transfer function of a section does not change by connecting it to the succeeding section in the cascade, and the overall transfer function of the cascade will be the product of the transfer functions of the individual sections.

To use the cascade design method, the transfer function $T(s)$ is first factored into the product of second-order functions and, for odd N, a first-order function. For example, consider the fifth-order low-pass filter whose transmission magnitude is shown in Fig. 14.3. Its transfer function in Eq. (14.8) can be factored as follows:

$$T(s) = \frac{k_1}{s+p_1} \times \frac{k_2(s^2+\omega_{l1}^2)}{s^2+b_{11}\,s+b_{01}} \times \frac{k_3(s^2+\omega_{l2}^2)}{s^2+b_{12}\,s+b_{02}}$$

where $k_1 k_2 k_3 = a_4$.

As a second example, consider the sixth-order bandpass filter whose transmission magnitude is shown in Fig. 14.4. Its transfer function in Eq. (14.9) can be factored as follows:

$$T(s) = \frac{k_1\, s}{s^2+b_{11}s+b_{01}} \times \frac{k_2(s^2+\omega_{l1}^2)}{s^2+b_{12}\,s+b_{02}} \times \frac{k_3(s^2+\omega_{l2}^2)}{s^2+b_{13}\,s+b_{03}}$$

In the remainder of this section we study first-order and second-order filter functions.

14.2.6 First-Order Filters

The general first-order transfer function is given by

$$T(s) = \frac{a_1\, s+a_0}{s+\omega_0} \tag{14.11}$$

This bilinear transfer function characterizes a first-order filter with a pole at $s = -\omega_0$, a transmission zero at $s = -a_0/a_1$, and a high-frequency gain that approaches a_1. The numerator coefficients, a_0 and a_1, determine the type of filter (e.g., low-pass, high-pass, etc.).

Example 14.2

Assuming that the op amp is ideal, show that the circuit in Fig. 14.8 realizes a first-order low-pass filter. Design the circuit to obtain a pole frequency $\omega_0 = 10^5$ rad/s and a dc-gain magnitude of 10 V/V. Use a 1 nF capacitor.

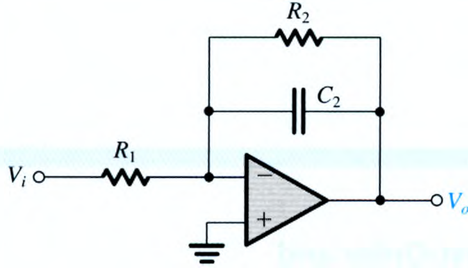


Figure 14.8 Circuit for Example 14.2.

Solution

$$\frac{V_o}{V_i} = -\frac{Z_2}{Z_1} = -\frac{Y_1}{Y_2} = -\frac{1/R_1}{\frac{1}{R_2} + s\,C_2}$$

Thus,

$$\frac{V_o}{V_i} - \frac{R_2}{R_1} = \frac{1}{1 + s\,C_2 R_2}$$

which is a first-order low-pass function with the pole-frequency ω_0 given by

$$\omega_0 = \frac{1}{C_2 R_2}$$

The dc gain is

$$\frac{V_o}{V_i}(s = 0) = -\frac{R_2}{R_1}$$

To obtain $\omega_0 = 10^5$ rad/s using $C_2 = 1$ nF, we select R_2 from

$$10^5 = \frac{1}{1 \times 10^{-9} \times R_2} \Rightarrow R_2 = 10^4\ \Omega = 10\ \text{k}\Omega$$

Finally, to obtain a dc gain of 10 V/V,

$$\frac{R_2}{R_1} = 10 \Rightarrow R_1 = \frac{R_2}{10} = \frac{10\ \text{k}\Omega}{10} = 1\ \text{k}\Omega$$

Figure H.1 (given in Appendix H on the website) provides a number of important special cases of the first-order filter function together with the corresponding passive (RC) and active (op amp–RC) realizations. Here we note that the active realizations, such as that in Example 14.2, are much more versatile than their passive counterparts; in many cases, the gain can be set to the desired value, and some transfer-function parameters can be adjusted without affecting others. Also, the output impedance of the active circuit is very low (ideally, zero), making cascading possible.

EXERCISES

14.6 For the low-pass filter designed in Example 14.2, assume that the maximum allowable variation in passband gain $A_{max} = 3$ dB. What is the passband-edge frequency, ω_p? If the stopband edge is $\omega_s = 3\,\omega_p$, find the minimum stopband attenuation, A_{min}. (*Hint*: Recall that A_{min} is measured relative to the maximum passband transmission.)

Ans. 10^5 rad/s; 10 dB

D14.7 Design the circuit of Fig. E14.7 to realize a high-pass filter with a 3-dB frequency of 10^4 rad/s and a high-frequency gain of 10 V/V. Use $R_1 = 10$ kΩ.

Ans. $R_2 = 100$ kΩ; $C_1 = 10$ nF

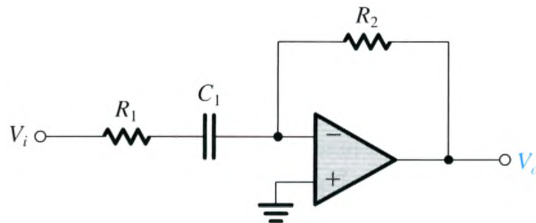


Figure E14.7

14.2.7 Second-Order Filter Functions

The general second-order or **biquadratic** transfer function is usually expressed in the standard form

$$T(s) = \frac{a_2 s^2 + a_1 s + a_0}{s^2 + s\left(\frac{\omega_0}{Q}\right) + \omega_0^2} \tag{14.12}$$

where ω_0 and Q determine the poles according to

$$p_1, p_2 = -\frac{\omega_0}{2Q} \pm j\omega_0\sqrt{1 - \frac{1}{4Q^2}} \tag{14.13}$$

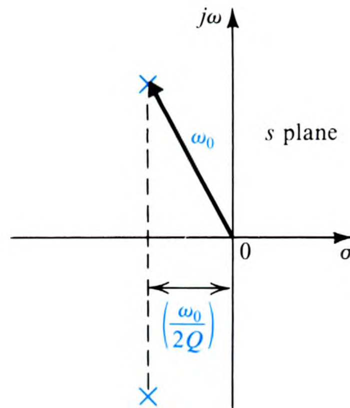


Figure 14.9 Definition of the parameters ω_0 and Q of a pair of complex-conjugate poles.

We are usually interested in the case of complex-conjugate poles, obtained for $Q > 0.5$. Figure 14.9 shows the location of the pair of complex-conjugate poles in the s plane. Notice that the radial distance of the poles from the origin is equal to ω_0, which is known as the **pole frequency**. The parameter Q determines the distance of the poles from the $j\omega$ axis: the higher the value of Q, the closer the poles are to the $j\omega$-axis, and the more selective the filter response becomes. An infinite value for Q locates the poles on the $j\omega$ axis and can yield sustained oscillations in the circuit realization. A negative value of Q implies that the poles are in the right half of the s plane, which certainly produces oscillations. The parameter Q is called the **pole quality factor**, or simply **pole** Q.

The transmission zeros of the second-order filter are determined by the numerator coefficients a_0, a_1, and a_2. It follows that the numerator coefficients determine the type of second-order filter function (i.e., LP, HP, etc.). Figure H.3 (provided in Appendix H, on the website) offers a complete tabulation of seven special second-order functions. For each case, we give the transfer function, the s-plane locations of the poles and zeros, and a detailed sketch of the magnitude response. Here, as examples, we consider three important special cases.

Low-Pass The second-order low-pass filter has the transfer function

$$T(s) = \frac{a_0}{s^2 + s\frac{\omega_0}{Q} + \omega_0^2} \tag{14.14}$$

which indicates that the two transmission zeros are at $s = \infty$. The magnitude response, obtained by setting $s = j\omega$, is shown in Fig. 14.10. As indicated, the response can exhibit a peak with the details shown. It can be shown that the peak occurs only for $Q > 1/\sqrt{2}$. The response obtained for $Q = 1/\sqrt{2}$ is called **maximally flat** (see the Butterworth filter in Section 14.3.1).

Bandpass The second-order bandpass filter has the transfer function

$$T(s) = \frac{a_1 s}{s^2 + s\frac{\omega_0}{Q} + \omega_0^2} \tag{14.15}$$

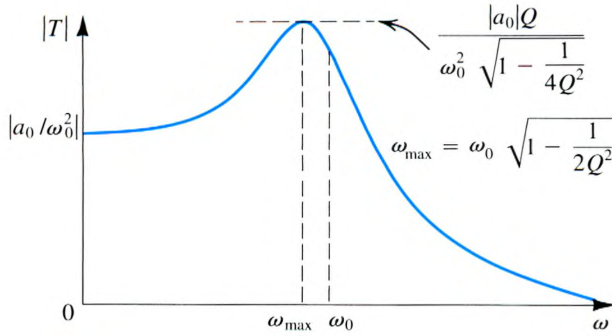


Figure 14.10 Magnitude response of a second-order low-pass filter.

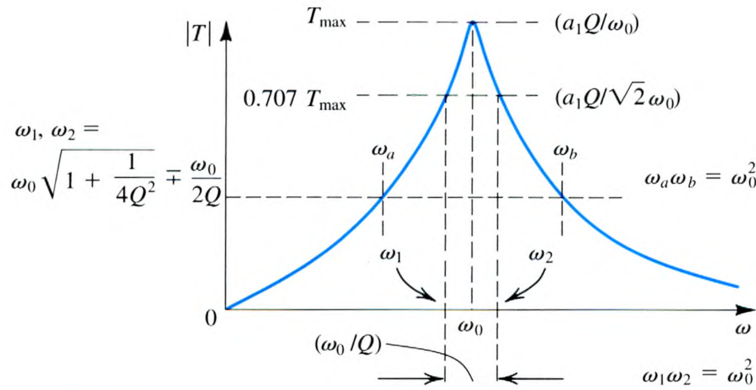


Figure 14.11 Magnitude response of a second-order bandpass filter.

which indicates that there is a transmission zero at $s = 0$ (dc) and the other transmission zero is at $s = \infty$. The magnitude function $|T(j\omega)|$, shown in Fig. 14.11, indicates that the response peaks at $\omega = \omega_0$. Thus, the **center frequency** of the bandpass is equal to the pole frequency ω_0. The selectivity of the second-order bandpass filter is usually measured by its *3-dB bandwidth*. This is the difference between the two frequencies ω_1 and ω_2, at which the magnitude response is 3 dB below its maximum value (at ω_0). It can be shown that

$$\omega_1, \omega_2 = \omega_0\sqrt{1+\frac{1}{4Q^2}} \mp \frac{\omega_0}{2Q} \tag{14.16}$$

Thus,

$$\text{BW} \equiv \omega_2 - \omega_1 = \frac{\omega_0}{Q} \tag{14.17}$$

Observe that as Q increases, the bandwidth decreases and the bandpass filter becomes more selective.

Notch The second-order notch filter has the transfer function

$$T(s) = a_2\frac{s^2 + \omega_n^2}{s^2 + s\frac{\omega_0}{Q} + \omega_0^2} \tag{14.18}$$

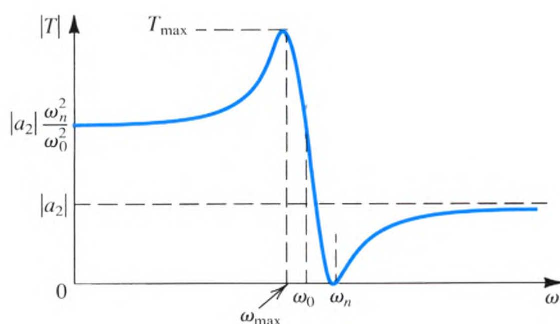


Figure 14.12 Magnitude response of a second-order low-pass notch (LPN) filter.

which indicates that the transmission zeros are at $s = \pm j\omega_n$. There are three special cases of the notch filter depending on the location of the notch frequency ω_n relative to the pole frequency ω_0. The magnitude response for the case $\omega_n > \omega_0$, known as a **low-pass notch (LPN)**, is shown in Fig. 14.12.

Most of the remainder of this chapter is devoted to the study of circuit realizations of second-order filter functions.

EXERCISES

14.8 For a second-order low-pass filter with $Q = 1/\sqrt{2}$ (the maximally flat case), show that at $\omega = \omega_0$ the magnitude response is 3 dB below the value at $\omega = 0$.

14.9 Give the transfer function of a second-order bandpass filter with a center frequency of 10^5 rad/s, a center-frequency gain of 10, and a 3-dB bandwidth of 10^3 rad/s.

Ans.

$$T(s) = \frac{10^4 s}{s^2 + 10^3 s + 10^{10}}$$

14.10 Give the transfer function of a low-pass notch filter with $\omega_0 = 1$ rad/s, $\omega_n = 1.2$ rad/s, $Q = 5$, and a dc gain of unity. What is the transmission at high frequencies?

Ans.

$$T(s) = 0.694\,\frac{s^2 + 1.44}{s^2 + 0.2\,s + 1}\ ;\ 0.694$$

14.3 Butterworth and Chebyshev Filters

In this section, we present two functions that are frequently used in approximating the transmission characteristics of low-pass filters; that is, in obtaining a transfer function $T(s)$ whose magnitude $|T(j\omega)|$ meets given low-pass filter specifications. Closed-form

expressions are available for the parameters of these functions, and thus one can use them in filter design without the need for computers or filter-design tables. Their utility, however, is limited to relatively simple applications.

Although in this section we discuss the design of low-pass filters only, the approximation functions presented can be applied to the design of other filter types through the use of frequency transformations (see Sedra and Brackett, 1978).

14.3.1 The Butterworth Filter

Figure 14.13 shows a sketch of the magnitude response of a Butterworth[2] filter. This filter exhibits a monotonically decreasing transmission with all the transmission zeros at $\omega = \infty$, making it an all-pole filter. The magnitude function for an *N*th-order Butterworth filter with a passband edge ω_p is given by

$$|T(j\omega)| = \frac{1}{\sqrt{1+\epsilon^2\left(\frac{\omega}{\omega_p}\right)^{2N}}} \tag{14.19}$$

At $\omega = \omega_p$,

$$|T(j\omega_p)| = \frac{1}{\sqrt{1+\epsilon^2}} \tag{14.20}$$

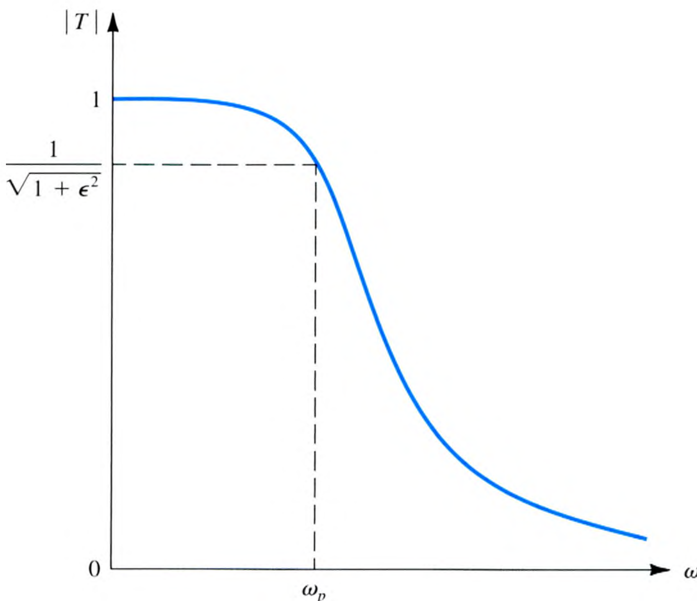


Figure 14.13 The magnitude response of a Butterworth filter.

[2]The Butterworth filter approximation is named after S. Butterworth, a British engineer who in 1930 was among the first to employ it.

Thus, the parameter ϵ determines the maximum variation in passband transmission, A_{max}, according to

$$A_{max} = 20 \log \sqrt{1+\epsilon^2} \tag{14.21}$$

Conversely, given A_{max}, the value of ϵ can be determined from

$$\epsilon = \sqrt{10^{A_{max}/10} - 1} \tag{14.22}$$

Observe that for the Butterworth response the maximum deviation in passband transmission (from the ideal value of unity) occurs at the passband edge only. It can be shown that the first $2N - 1$ derivatives of $|T|$ relative to ω are zero at $\omega = 0$ (see Van Valkenburg, 1980). This property makes the Butterworth response very flat near $\omega = 0$ and results in the name **maximally flat** response. The degree of passband flatness increases as the order N is increased, as can be seen from Fig. 14.14. This figure indicates also that, as should be expected, as the order N is increased, the filter response approaches the ideal brick-wall type of response. Finally, note from Eq. (14.21) that for $\epsilon = 1$, $A_{max} = 3$ dB and ω_p is the 3-dB bandwidth.

At the edge of the stopband, $\omega = \omega_s$, the attenuation of the Butterworth filter can be obtained by substituting $\omega = \omega_s$ in Eq. (14.19). The result is given by

$$A(\omega_s) = -20\log\left[1\Big/\sqrt{1+\epsilon^2(\omega_s/\omega_p)^{2N}}\right] \tag{14.23}$$

$$= 10\log\left[1+\epsilon^2(\omega_s/\omega_p)^{2N}\right]$$

This equation can be used to determine the filter order required, which is the lowest integer value of N that yields $A(\omega_s) \geq A_{min}$.

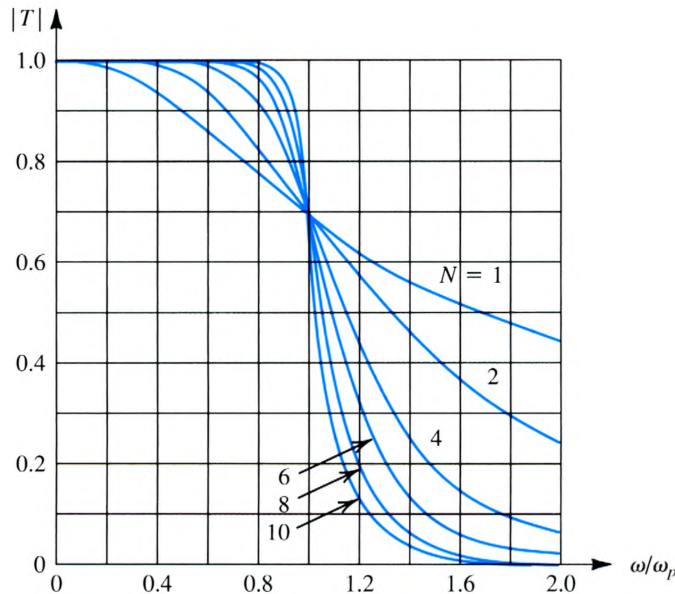


Figure 14.14 Magnitude response for Butterworth filters of various order with $\epsilon = 1$. Note that as the order increases, the response approaches the ideal brick-wall type of transmission.

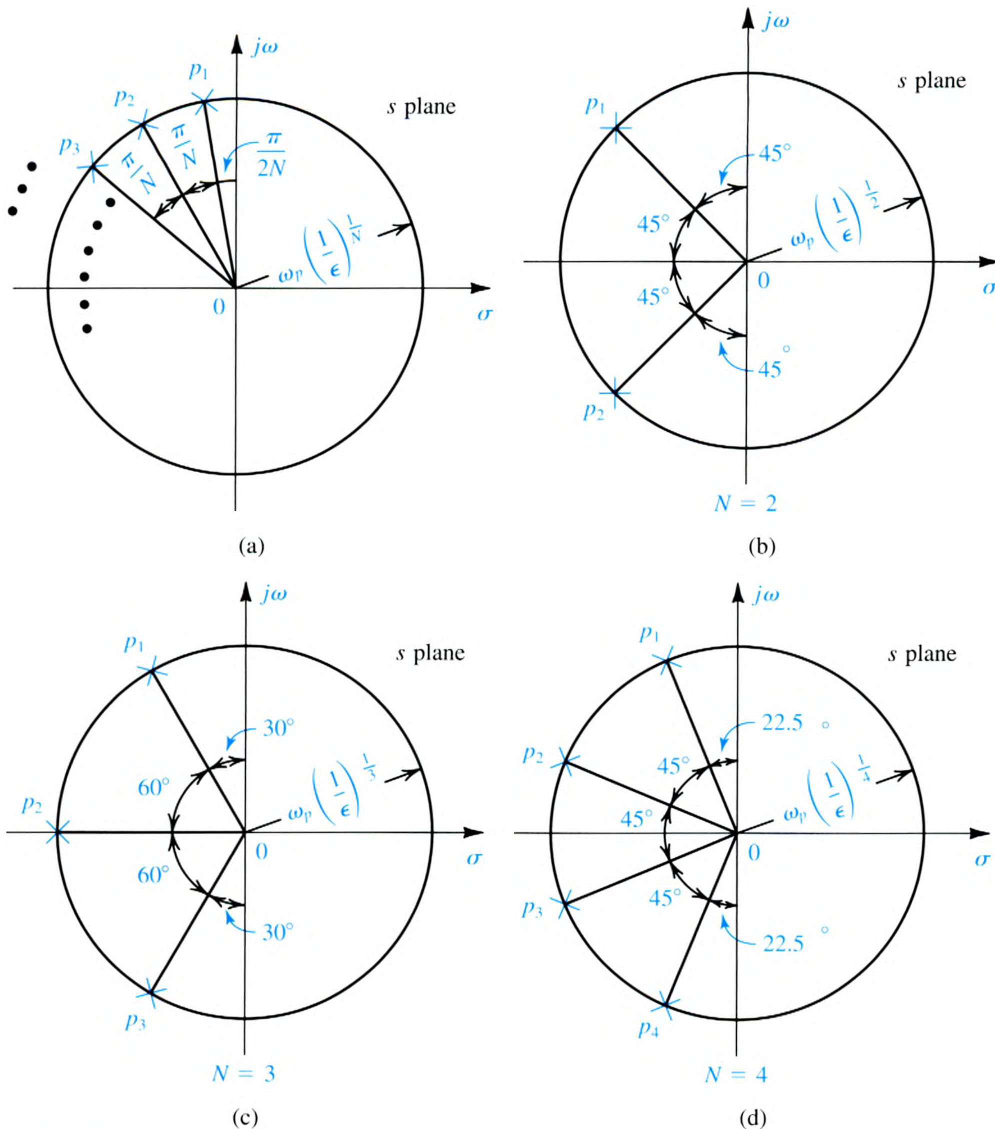


Figure 14.15 Graphical construction for determining the poles of a Butterworth filter of order N. All the poles lie in the left half of the s plane on a circle of radius $\omega_0 = \omega_p(1/\epsilon)^{1/N}$, where ϵ is the passband deviation parameter $\left(\epsilon = \sqrt{10^{A_{\max}/10} - 1}\right)$: **(a)** the general case; **(b)** $N = 2$; **(c)** $N = 3$; **(d)** $N = 4$.

The poles of an Nth-order Butterworth filter[3] can be determined from the graphical construction shown in Fig. 14.15(a). Observe that the poles lie on a circle of radius $\omega_p(1/\epsilon)^{1/N}$ and are spaced by equal angles of π/N, with the first pole at an angle $\pi/2N$ from the $+j\omega$ axis. Figures 14.15(b), (c), and (d) show the poles for the cases $N = 2$, 3, and 4, respectively. As

[3]Here we provide the pole locations without derivation. The derivation can be found in one of the textbooks on filters, such as Sedra and Brackett (1978) or Schaumann et al. (2010).

expected, the poles occur in complex-conjugate pairs with a real axis pole at $s = -\omega_p(1/\epsilon)^{1/N}$ occurring for the case N is odd.

Since the poles all have equal radial distances from the origin, they all have the same pole frequency:

$$\omega_0 = \omega_p(1/\epsilon)^{1/N} \tag{14.24}$$

We can use the graphical construction of Fig. 14.15(a) to derive a simple expression for the Q factor for each complex-conjugate pair of poles. Noting from Fig. 14.9 that the horizontal distance of a pair of complex-conjugate poles from the $j\omega$-axis is $(\omega_0/2Q)$, we can write for (p_1, p_1^*)

$$\omega_0 \sin\left(\frac{\pi}{2N}\right) = \frac{\omega_0}{2Q_1}$$

which yields

$$Q_1 = \frac{1}{2\sin(\pi/2N)}$$

Generally, for the pole pair (p_k, p_k^*),

$$Q_k = 1\Big/\left[2\sin\left(\frac{2k-1}{N}\frac{\pi}{2}\right)\right] \tag{14.25}$$

where

$$k = 1, 2, \ldots, \left(\frac{N-1}{2}\right) \text{ for } N \text{ odd}$$

and

$$k = 1, 2, \ldots, \frac{N}{2} \text{ for } N \text{ even}$$

Once we have determined the poles, we can write the transfer function $T(s)$ of the Nth order Butterworth filter as

$$T(s) = \frac{K\,\omega_0^N}{(s+\omega_0)\displaystyle\prod_{k=1}^{(N-1)/2}\left(s^2 + s\frac{\omega_0}{Q_k} + \omega_0^2\right)}, \quad \text{for } N \text{ odd} \tag{14.26a}$$

or

$$T(s) = \frac{K\,\omega_0^N}{\displaystyle\prod_{k=1}^{N/2}\left(s^2 + s\frac{\omega_0}{Q_k} + \omega_0^2\right)}, \quad \text{for } N \text{ even} \tag{14.26b}$$

where K is a constant equal to the required dc gain of the filter.

To summarize, to find a Butterworth transfer function that meets transmission specifications of the form in Fig. 14.3, we perform the following procedure:

1. Determine ϵ from Eq. (14.22).
2. Use Eq. (14.23) to determine the required filter order as the lowest integer value of N that results in $A(\omega_s) \geq A_{\min}$.
3. Use Fig. 14.15(a) together with Eqs. (14.24) and (14.25) to determine the poles.
4. Use Eq. (14.26a) or Eq. (14.26b) to determine $T(s)$.

Example 14.3

Find the Butterworth transfer function that meets the following low-pass filter specifications: $f_p = 10$ kHz, $A_{max} = 1$ dB, $f_s = 15$ kHz, $A_{min} = 25$ dB, and dc gain = 1.

Solution

Substituting $A_{max} = 1$ dB into Eq. (14.22) yields $\epsilon = 0.5088$. We can then use Eq. (14.23) to determine the required filter order by trying various values for N. We find that $N = 8$ yields $A(\omega_s) = 22.3$ dB and $N = 9$ yields $A(\omega_s) = 25.8$ dB. So, we select $N = 9$.

Figure 14.16 shows the graphical construct for determining the poles. The frequency ω_0 for the 9 poles can be found using Eq. (14.24):

$$\omega_0 = 2\pi \times 10 \times 10^3 (1/0.5088)^{1/9} = 6.773 \times 10^4 \text{ rad/s},$$

The Q factors for the 4 pairs of complex-conjugate poles can be obtained using Eq. (14.25):

$$Q_1 = 2.879, \; Q_2 = 1, \; Q_3 = 0.653, \; Q_4 = 0.532$$

Finally, $T(s)$ can be obtained by substituting in Eq. (14.26a) with $K = 1$:

$$T(s) = \frac{\omega_0^9}{(s+\omega_0)\left(s^2 + s\frac{\omega_0}{2.879} + \omega_0^2\right)\left(s^2 + s\omega_0 + \omega_0^2\right)} \times \frac{1}{\left(s^2 + s\frac{\omega_0}{0.653} + \omega_0^2\right)\left(s^2 + s\frac{\omega_0}{0.532} + \omega_0^2\right)} \quad (14.27)$$

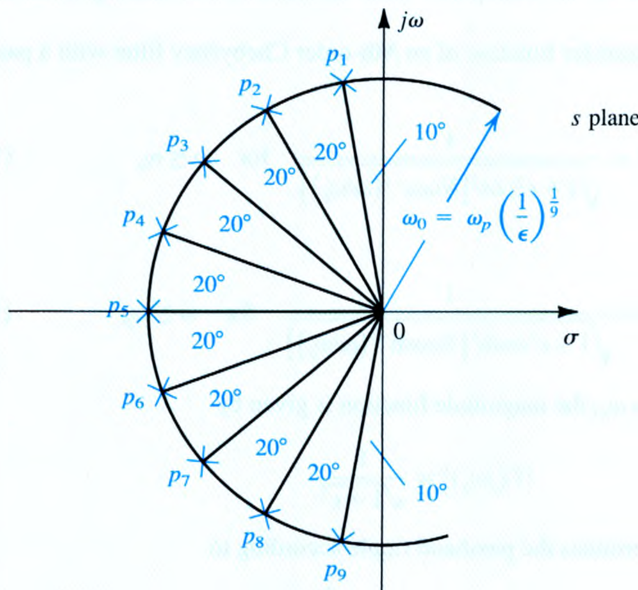


Figure 14.16 Poles of the ninth-order Butterworth filter of Example 14.3.

EXERCISES

14.11 For a fifth-order Butterworth low-pass filter with dc transmission of unity and 3-dB bandwidth of 10 kHz, find the attenuation obtained at 30 kHz.
Ans. 47.7 dB

D14.12 Determine the order N of a Butterworth filter for which $A_{max} = 1$ dB, $\omega_s/\omega_p = 1.5$, and $A_{min} = 30$ dB.
Ans. $N = 11$

14.13 Find the poles and the transfer function of a third-order Butterworth filter with $\omega_p = 1$ rad/s, $A_{max} =$ 3 dB ($\epsilon \simeq 1$), and the dc gain is unity.
Ans. Poles: a real pole at $s = -1$ and a pair of complex-conjugate poles with $\omega_0 = 1$ rad/s and $Q = 1$; $T(s) = 1/[(s+1)(s^2+s+1)]$.

14.3.2 The Chebyshev Filter

Figure 14.17 shows representative transmission functions for Chebyshev[4] filters of even and odd orders. The Chebyshev filter exhibits an equiripple response in the passband and a monotonically decreasing transmission in the stopband. While the odd-order filter has $|T(0)| = 1$, the even-order filter exhibits its maximum magnitude deviation at $\omega = 0$. In both cases the total number of passband maxima and minima equals the order of the filter, N. All the transmission zeros of the Chebyshev filter are at $\omega = \infty$, making it an all-pole filter.

The magnitude of the transfer function of an Nth-order Chebyshev filter with a passband edge, ω_p is given by

$$|T(j\omega)| = \frac{1}{\sqrt{1+\epsilon^2\cos^2\left[N\cos^{-1}\left(\omega/\omega_p\right)\right]}} \quad \text{for} \quad \omega \le \omega_p \tag{14.28}$$

and

$$|T(j\omega)| = \frac{1}{\sqrt{1+\epsilon^2\cosh^2\left[N\cosh^{-1}\left(\omega/\omega_p\right)\right]}} \quad \text{for} \quad \omega \ge \omega_p \tag{14.29}$$

At the passband edge, $\omega = \omega_p$, the magnitude function is given by

$$\left|T\left(j\omega_p\right)\right| = \frac{1}{\sqrt{1+\epsilon^2}}$$

Thus, the parameter ϵ determines the passband ripple according to

$$A_{max} = 10\log\left(1+\epsilon^2\right) \tag{14.30}$$

[4]Named after the Russian mathematician P. L. Chebyshev, who in 1899 used these functions in studying the construction of steam engines.

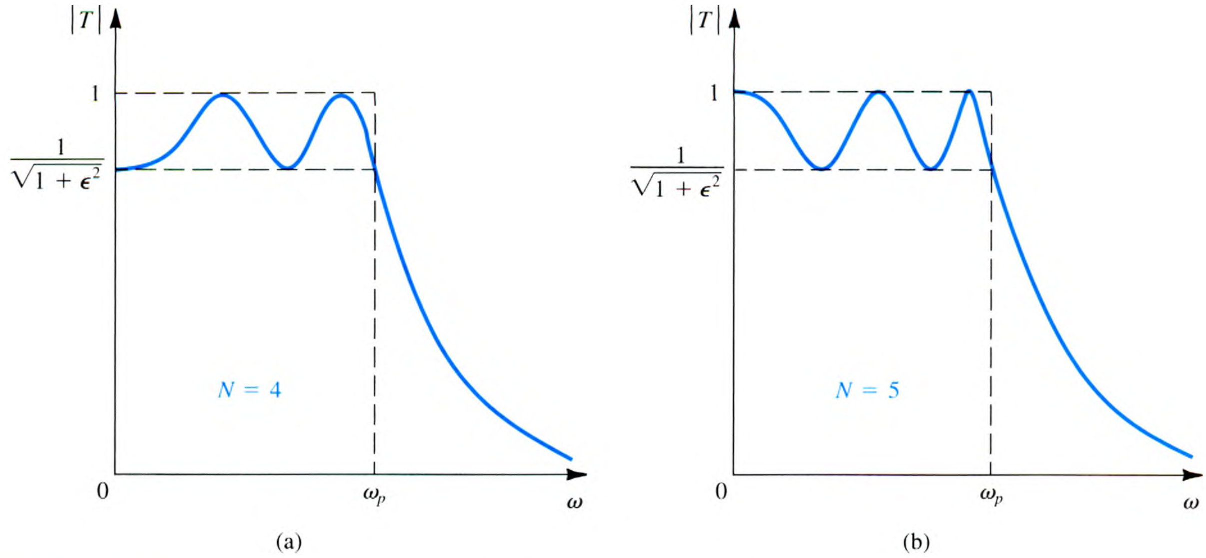


Figure 14.17 Sketches of the transmission characteristics of representative **(a)** even-order and **(b)** odd-order Chebyshev filters.

Conversely, given A_{max}, the value of ϵ is determined from

$$\epsilon = \sqrt{10^{A_{max}/10} - 1} \tag{14.31}$$

The attenuation achieved by the Chebyshev filter at the stopband edge ($\omega = \omega_s$) is found using Eq. (14.29) as

$$A(\omega_s) = 10\log\left[1 + \epsilon^2 \cosh^2\left(N\cosh^{-1}(\omega_s/\omega_p)\right)\right] \tag{14.32}$$

With the aid of a calculator, this equation can be used to determine the order N required to obtain a specified A_{min} by finding the lowest integer value of N that yields $A(\omega_s) \geq A_{min}$. As in the case of the Butterworth filter, increasing the order N of the Chebyshev filter causes its magnitude function to approach the ideal brick-wall low-pass response.

The poles of the Chebyshev filter are given by

$$p_k = -\omega_p \sin\left(\frac{2k-1}{N}\frac{\pi}{2}\right)\sinh\left(\frac{1}{N}\sinh^{-1}\frac{1}{\epsilon}\right) \tag{14.33}$$

$$+j\omega_p \cos\left(\frac{2k-1}{N}\frac{\pi}{2}\right)\cosh\left(\frac{1}{N}\sinh^{-1}\frac{1}{\epsilon}\right) \qquad k = 1, 2, \ldots, N$$

The frequency ω_0 and the Q factor for each complex-conjugate pair of poles can be determined from this equation together with Fig. 14.9.

Finally, the transfer function of the Chebyshev filter can be written as

$$T(s) = \frac{K\omega_p^N}{\epsilon 2^{N-1}(s-p_1)(s-p_2)\cdots(s-p_N)} \tag{14.34}$$

where K is the dc gain that the filter is required to have.

To summarize, given low-pass transmission specifications of the type shown in Fig. 14.3, the transfer function of a Chebyshev filter that meets these specifications can be found as follows:

1. Determine ϵ from Eq. (14.31).
2. Use Eq. (14.32) to determine the order required.
3. Determine the poles using Eq. (14.33).
4. Determine the transfer function using Eq. (14.34).

The Chebyshev filter provides a more efficient approximation than the Butterworth filter. Thus, for the same order and the same A_{max}, the Chebyshev filter provides greater stopband attenuation than the Butterworth filter. Alternatively, to meet identical specifications, one requires a lower order for the Chebyshev than for the Butterworth filter. The price paid, however, is higher values for the Q factors of the poles. In later sections, we will see that poles with high Q factors are more difficult to realize.

Example 14.4

Find the Chebyshev transfer function that meets the same low-pass filter specifications given in Example 14.3: namely, $f_p = 10$ kHz, $A_{max} = 1$ dB, $f_s = 15$ kHz, $A_{min} = 25$ dB, dc gain = 1.

Solution

Substituting $A_{max} = 1$ dB into Eq. (14.31) yields $\epsilon = 0.5088$. By trying various values for N in Eq. (14.32) we find that $N = 4$ yields $A(\omega_s) = 21.6$ dB and $N = 5$ provides 29.9 dB. We thus select $N = 5$. Recall that we required a ninth-order Butterworth filter to meet the same specifications in Example 14.3.

The poles are obtained by substituting in Eq. (14.33) as

$$p_1, p_5 = \omega_p(-0.0895 \pm j0.9901) \Rightarrow \omega_{01} = 0.9941\,\omega_p, Q_1 = 5.554$$

$$p_2, p_4 = \omega_p(-0.2342 \pm j0.6119) \Rightarrow \omega_{02} = 0.6552\,\omega_p, Q_2 = 1.4$$

$$p_5 = \omega_p(-0.2895)$$

The transfer function is obtained by substituting these values in Eq. (14.34) as

$$T(s) = \frac{\omega_p^5}{8.1408(s + 0.2895\,\omega_p)(s^2 + s\frac{0.9941\omega_p}{5.554} + 0.9882\,\omega_p^2)} \times \frac{1}{s^2 + s\,\frac{0.6552\omega_p}{1.4} + 0.4293\,\omega_p^2} \tag{14.35}$$

where $\omega_p = 2\pi \times 10^4$ rad/s.

EXERCISES

D14.14 Find the attenuation provided at $\omega = 2\omega_p$ by a seventh-order Chebyshev filter with a 0.5-dB passband ripple. If the passband ripple is allowed to increase to 1 dB, by how much does the stopband attenuation increase?
Ans. 64.9 dB; 3.3 dB

D14.15 It is required to design a low-pass filter having $f_p = 1$ kHz, $A_{max} = 1$ dB, $f_s = 1.5$ kHz, $A_{min} =$ 50 dB. (a) Find the required order of a Chebyshev filter. What is the excess stopband attenuation obtained? (b) Repeat for a Butterworth filter.
Ans. (a) $N = 8$, 5 dB; (b) $N = 16$, 0.5 dB

EARLY FILTER PIONEERS: CAUER AND DARLINGTON

While on a fellowship with Vannevar Bush at MIT and Harvard, the German mathematician Wilhelm Cauer (1900–1945) used the Chebyshev polynomials in a way that unified the field of filter transfer function design. The elliptical filters now known as Cauer filters have equiripple performance in both the passband and the stopband(s). Cauer continued to make contributions to LC filter synthesis until his mysterious disappearance and presumed death in Berlin on the last day of the Second World War.

Sidney Darlington (1906–1997) developed a complete design theory for LC filters while working at the Bell Telephone Laboratories in the 1940s. Ironically, in later years he became better known for his invention of a particular transistor circuit, the Darlington pair.

14.4 Second-Order Passive Filters Based on the LCR Resonator

In this section we study the realization of second-order filter functions using inductors, capacitors, and resistors. These LCR circuits are useful in their own right for filters operating at very high frequencies. As well, they serve as the basis for the op amp–RC realizations presented in the next section, where the inductor is replaced with a simulated inductance.

The LCR realizations studied in this section are derived from the parallel LCR resonance circuit shown in Fig. 14.18(a).

14.4.1 The Resonator Poles

The poles of the parallel resonance circuit of Fig. 14.18(a) can be determined by applying *an excitation that does not change the natural structure of the circuit.* Two possible ways of exciting the circuit are shown in Figs. 14.18(b) and (c). In Fig. 14.18(b) the resonator is excited with a current source I connected in parallel. Since, as far as the natural response of a

circuit is concerned, an independent ideal current source is equivalent to an open circuit, the excitation of Fig. 14.18(b) does not alter the natural structure of the resonator. Thus the circuit in Fig. 14.18(b) can be used to determine the poles of the resonator by simply finding the poles of any response function. We can for instance take the voltage V_o across the resonator as the response and thus obtain the response function $V_o/I = Z$, where Z is the impedance of the parallel resonance circuit. However, because of the parallel structure of the circuit it is more convenient to work in terms of the admittance Y; thus,

$$\frac{V_o}{I} = \frac{1}{Y} = \frac{1}{(1/sL) + sC + (1/R)} \tag{14.36}$$

$$= \frac{s/C}{s^2 + s(1/CR) + (1/LC)}$$

Equating the denominator to the standard form $[s^2 + s(\omega_0/Q) + \omega_0^2]$ leads to

$$\omega_0^2 = 1/LC \tag{14.37}$$

and

$$\omega_0/Q = 1/CR \tag{14.38}$$

Thus,

$$\omega_0 = 1/\sqrt{LC} \tag{14.39}$$

$$Q = \omega_0 CR \tag{14.40}$$

These expressions should be familiar to you from studies of parallel resonance circuits in introductory courses on circuit analysis.

An alternative way of exciting the parallel LCR resonator for the purpose of determining its poles is shown in Fig. 14.18(c). Here, node x of inductor L has been disconnected from ground and connected to an ideal voltage source V_i. Now, since as far as the natural response of a circuit is concerned, an ideal independent voltage source is equivalent to a short circuit, the excitation of Fig. 14.18(c) does not alter the natural structure of the resonator. Thus we can use the circuit in Fig. 14.18(c) to determine the poles of the resonator. These are the poles of any response function. For instance, we can select V_o as the response variable and find the

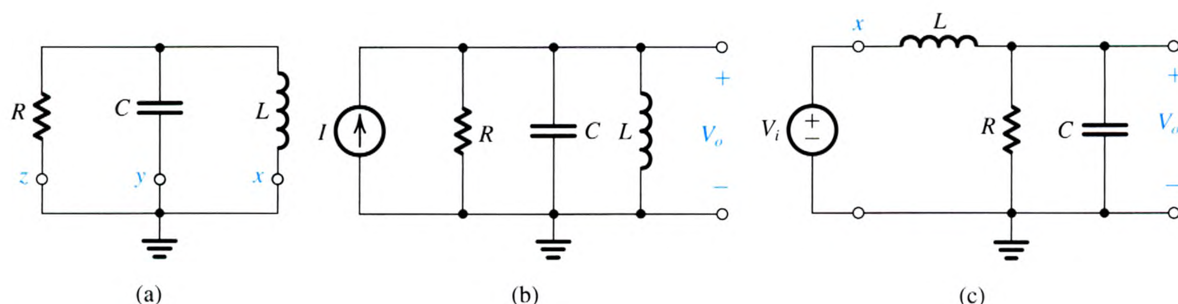


Figure 14.18 **(a)** The second-order parallel LCR resonator. **(b, c)** Two ways of exciting the resonator of **(a)** without changing its *natural structure*; resonator poles are those poles of V_o/I and V_o/V_i.

transfer function V_o/V_i. You can easily verify that this will lead to the poles determined earlier.

In a design problem, we will be given ω_0 and Q, and will be asked to determine L, C, and R. Equations (14.39) and (14.40) are two equations in the three unknowns. The one available degree of freedom can be utilized to set the impedance level of the circuit to a value that results in practical component values.

14.4.2 Realization of Transmission Zeros

Having selected the component values of the LCR resonator to realize a given pair of complex-conjugate poles, we now consider the use of the resonator to realize a desired filter type (e.g., LP, HP, etc.). Specifically, we wish to find out where to inject the input voltage signal V_i so that the transfer function V_o/V_i is the desired one. Toward that end, note that in the resonator circuit in Fig. 14.18(a), any of the nodes labeled x, y, or z can be disconnected from ground and connected to V_i without altering the circuit's natural structure and thus its poles. When this is done, the circuit takes the form of a voltage divider, as shown in Fig. 14.19(a). Thus the transfer function realized is

$$T(s) = \frac{V_o(s)}{V_i(s)} = \frac{Z_2(s)}{Z_1(s) + Z_2(s)} \tag{14.41}$$

We observe that *the transmission zeros are the values of s at which $Z_2(s)$ is zero, provided $Z_1(s)$ is not simultaneously zero, and the values of s at which $Z_1(s)$ is infinite, provided $Z_2(s)$ is not simultaneously infinite*. This statement makes physical sense: The output will be zero either when $Z_2(s)$ behaves as a short circuit or when $Z_1(s)$ behaves as an open circuit. If there is a value of s at which both Z_1 and Z_2 are zero, then V_o/V_i will be finite and no transmission zero is obtained. Similarly, if there is a value of s at which both Z_1 and Z_2 are infinite, then V_o/V_i will be finite and no transmission zero is realized. In the following, we present three examples of LCR filter realizations.

14.4.3 Realization of the Low-Pass Function

Using the scheme just outlined, we see that to realize a low-pass function, node x is disconnected from ground and connected to V_i, as shown in Fig. 14.19(b). The transmission zeros of this circuit will be at the value of s for which the series impedance becomes infinite (sL becomes infinite at $s = \infty$) and the value of s at which the shunt impedance becomes zero ($1/[sC + (1/R)]$ becomes zero at $s = \infty$). Thus this circuit has two transmission zeros at $s = \infty$, as a second-order LP is supposed to. The transfer function can be written either by inspection or by using the voltage divider rule. The latter approach gives us

$$T(s) \equiv \frac{V_o}{V_i} = \frac{Z_2}{Z_1 + Z_2} = \frac{Y_1}{Y_1 + Y_2} = \frac{1/sL}{(1/sL) + sC + (1/R)}$$

$$= \frac{1/LC}{s^2 + s(1/CR) + (1/LC)} \tag{14.42}$$

14.4.4 Realization of the Bandpass Function

The bandpass function is realized by disconnecting node z from ground and connecting it to V_i, as shown in Fig. 14.19(c). Here the series impedance is resistive and thus does

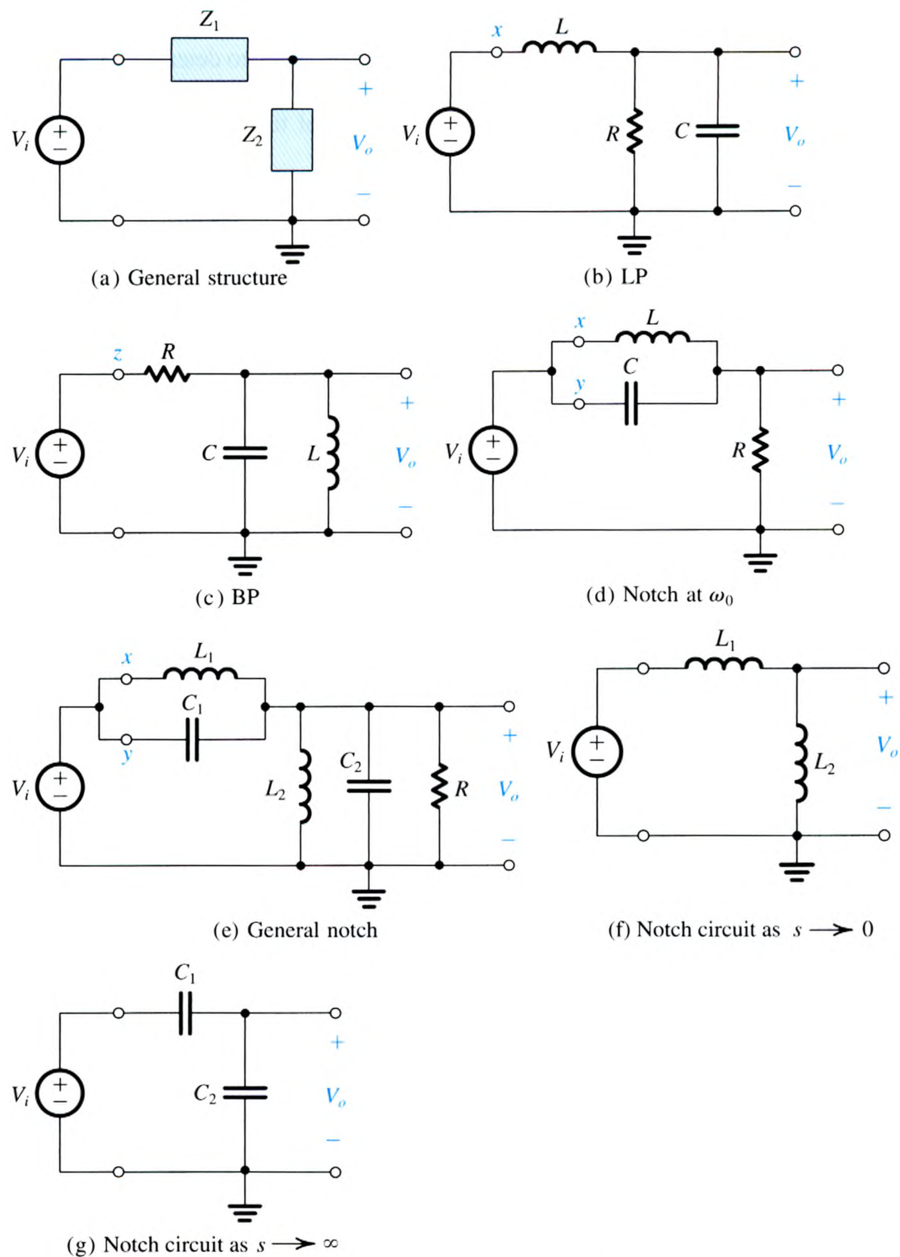


Figure 14.19 Realization of various second-order filter functions using the LCR resonator of Fig. 14.18(a): **(a)** general structure, **(b)** LP, **(c)** BP, **(d)** notch at ω_0, **(e)** general notch, **(f)** notch circuit in (e) as $s \to 0$, **(g)** notch circuit in (e) as $s \to \infty$.

not introduce any transmission zeros. These are obtained as follows: One zero at $s = 0$ is realized by the shunt inductor, and one zero at $s = \infty$ is realized by the shunt capacitor. At the center frequency ω_0, the parallel LC-tuned circuit exhibits an infinite impedance, and thus no current flows through R. It follows that at $\omega = \omega_0$, $V_o = V_i$. In other words, the

center-frequency gain of the bandpass filter is unity. Its transfer function can be obtained as follows:

$$T(s) = \frac{Y_R}{Y_R + Y_L + Y_C} = \frac{1/R}{(1/R) + (1/sL) + sC}$$
$$= \frac{s(1/CR)}{s^2 + s(1/CR) + (1/LC)} \qquad (14.43)$$

14.4.5 Realization of the Notch Functions

To obtain a pair of transmission zeros on the $j\omega$ axis, we use a parallel resonance circuit in the series arm, as shown in Fig. 14.19(d). Observe that this circuit is obtained by disconnecting both nodes x and y from ground and connecting them together to V_i. The impedance of the LC circuit becomes infinite at $\omega = \omega_0 = 1/\sqrt{LC}$, thus causing zero transmission at this frequency. The shunt impedance is resistive and thus does not introduce transmission zeros. It follows that the circuit in Fig. 14.19(d) will realize the notch transfer function

$$T(s) = a_2 \frac{s^2 + \omega_0^2}{s^2 + s(\omega_0/Q) + \omega_0^2} \qquad (14.44)$$

The value of the high-frequency gain a_2 can be found from the circuit to be unity.

To obtain a notch-filter realization in which the notch frequency ω_n is arbitrarily placed relative to ω_0, we adopt a variation on the scheme above. We still use a parallel LC circuit in the series branch, as shown in Fig. 14.19(e), where L_1 and C_1 are selected so that

$$L_1 C_1 = 1/\omega_n^2 \qquad (14.45)$$

Thus the $L_1 C_1$ tank circuit will introduce a pair of transmission zeros at $\pm j\omega_n$, provided the $L_2 C_2$ tank is not resonant at ω_n. Apart from this restriction, the values of L_2 and C_2 must be selected to ensure that the poles have not been altered; thus,

$$C_1 + C_2 = C \qquad (14.46)$$

$$L_1 \| L_2 = L \qquad (14.47)$$

In other words, when V_i is replaced by a short circuit, the circuit should reduce to the original LCR resonator. Another way of thinking about the circuit of Fig. 14.19(e) is that it is obtained from the original LCR resonator by lifting part of L and part of C off ground and connecting them to V_i.

It should be noted that in the circuit of Fig. 14.19(e), L_2 does *not* introduce a zero at $s = 0$ because at $s = 0$, the $L_1 C_1$ circuit also has a zero. In fact, at $s = 0$ the circuit reduces to the inductive voltage divider shown in Fig. 14.19(f) with the dc transmission being $L_2/(L_1 + L_2)$. Similar comments can be made about C_2 and the fact that it does *not* introduce a zero at $s = \infty$ and that the transmission at $s = \infty$ is $C_1/(C_1 + C_2)$ [see Fig. 14.19(g)].

The LPN and HPN filter realizations are special cases of the general notch circuit of Fig. 14.19(f) and are explored in Problems 14.48 and 14.49, respectively.

EXERCISES

D14.16 Use the circuit of Fig. 14.19(b) to realize a second-order low-pass function of the maximally flat type with a 3-dB frequency of 100 kHz.
Ans. Selecting $R = 1$ kΩ, we obtain $C = 1125$ pF and $L = 2.25$ mH.

D14.17 Use the circuit in Fig. 14.19(c) to design a bandpass filter with a center-frequency of 10^5 rad/s and a 3-dB bandwidth of 10^3 rad/s.
Ans. Selecting $R = 100$ kΩ we obtain $C = 10$ nF and $L = 10$ mH.

14.5 Second-Order Active Filters Based on Inductance Simulation

In this section, we study a family of op amp–RC circuits that realize the various second-order filter functions. The circuits are based on an op amp–RC resonator obtained by replacing the inductor L in the LCR resonator with an op amp–RC circuit that has an inductive input impedance.

14.5.1 The Antoniou Inductance-Simulation Circuit

Over the years, many op amp–RC circuits have been proposed for simulating the operation of an inductor. Of these, one circuit, invented by A. Antoniou[5] (see Antoniou, 1969), has proved to be the "best." By "best" we mean that the operation of the circuit is very tolerant of the nonideal properties of the op amps, in particular their finite gain and bandwidth. Figure 14.20(a) shows the Antoniou inductance-simulation circuit. If the circuit is fed at its input (node 1) with a voltage source V_1 and the input current is denoted I_1, then for ideal op amps the input impedance can be shown to be

$$Z_{in} \equiv V_1/I_1 = sC_4R_1R_3R_5/R_2 \tag{14.48}$$

which is that of an inductance L given by

$$L = C_4R_1R_3R_5/R_2 \tag{14.49}$$

Figure 14.20(b) shows the analysis of the circuit assuming that the op amps are ideal and thus that a virtual short circuit appears between the two input terminals of each op amp, and assuming also that the input currents of the op amps are zero. The analysis begins at node 1, which is assumed to be fed by a voltage source V_1, and proceeds step by step, with the order of the steps indicated by the circled numbers. The result of the analysis is the expression shown for the input current I_1 from which Z_{in} is found.

[5]Andreas Antoniou is a Canadian academic, currently (2019) a professor emeritus of electrical and computer engineering at the University of Victoria, Victoria, British Columbia.

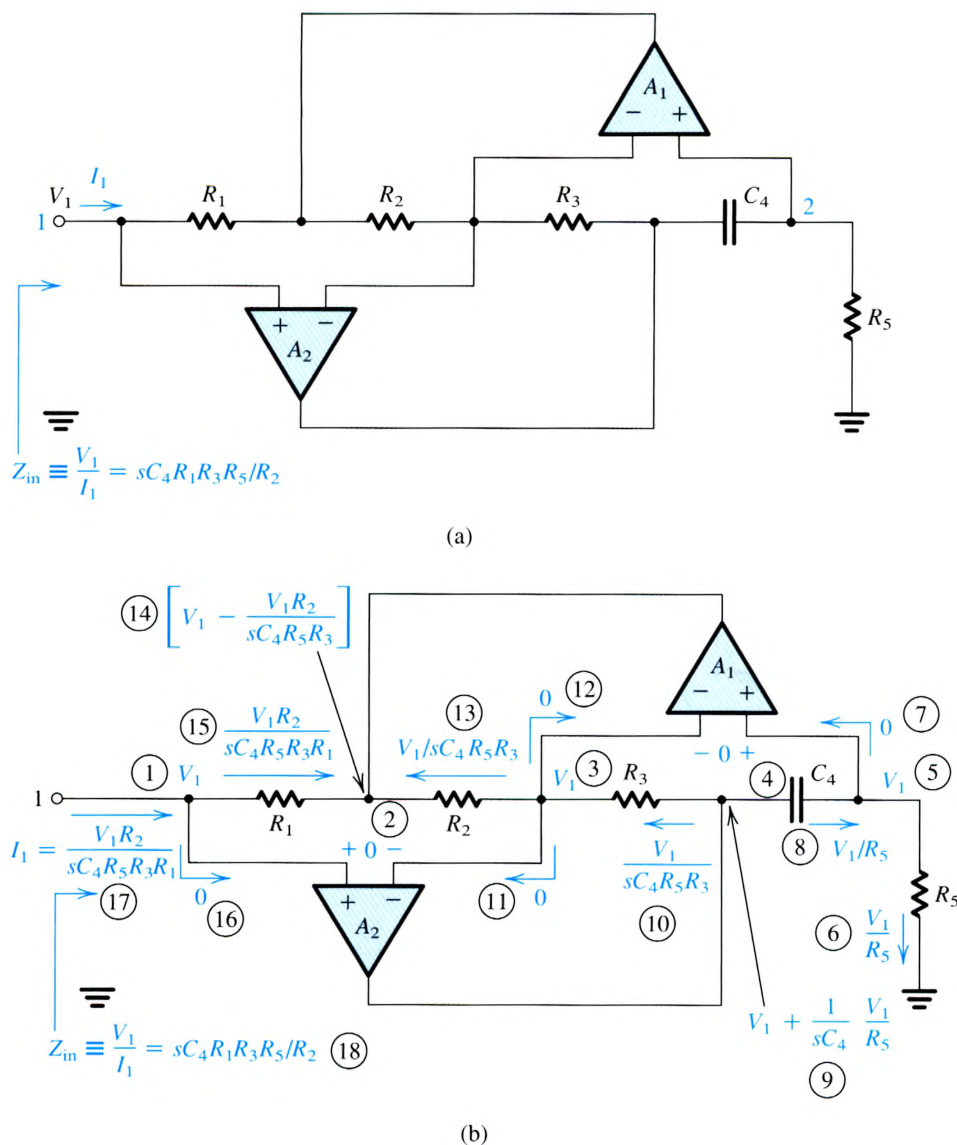


Figure 14.20 **(a)** The Antoniou inductance-simulation circuit. **(b)** Analysis of the circuit assuming ideal op amps. The order of the analysis steps is indicated by the circled numbers.

The design of this circuit is usually based on selecting $R_1 = R_2 = R_3 = R_5 = R$ and $C_4 = C$, which leads to $L = CR^2$. Convenient values are then selected for C and R to yield the desired inductance value L. More details on this circuit and the effect of the nonidealities of the op amps on its performance can be found in Sedra and Brackett (1978).

14.5.2 The Op Amp–RC Resonator

Figure 14.21(a) shows the LCR resonator we studied in detail in Section 14.4. Replacing the inductor L with a simulated inductance realized by the Antoniou circuit of Fig. 14.20(a)

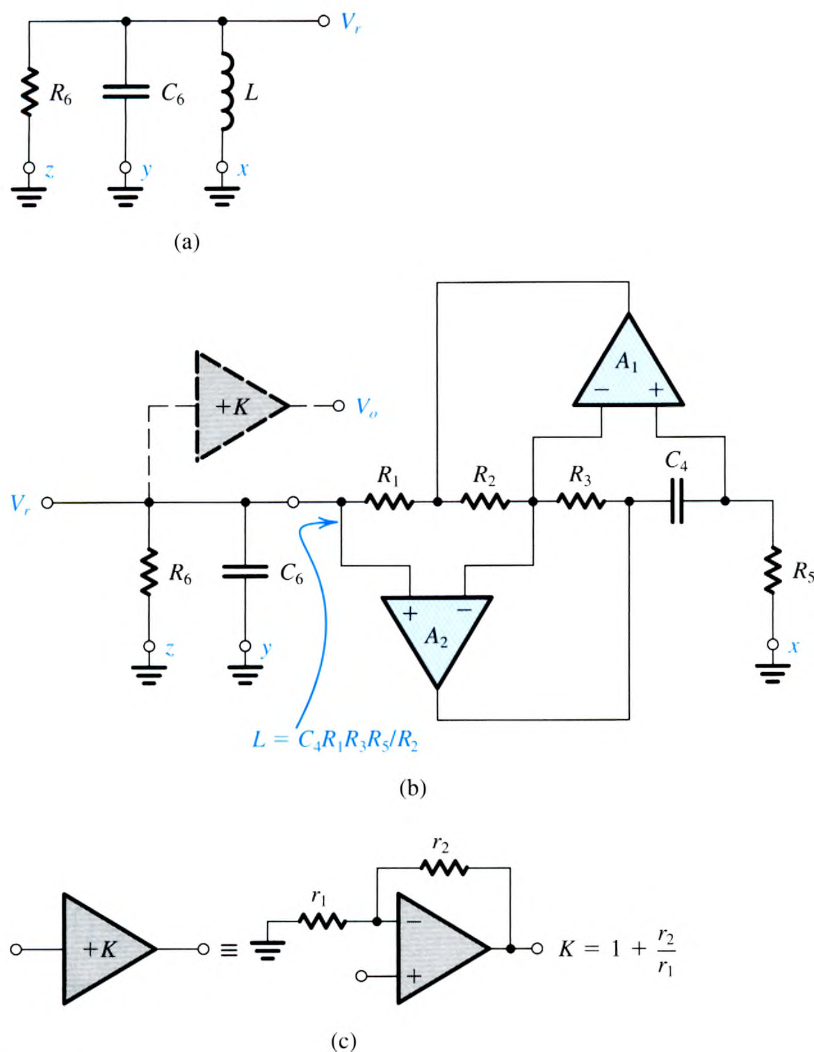


Figure 14.21 **(a)** An LCR resonator. **(b)** An op amp–RC resonator obtained by replacing the inductor L in the LCR resonator of **(a)** with a simulated inductance realized by the Antoniou circuit of Fig. 14.20(a). **(c)** Implementation of the buffer amplifier K.

results in the op amp–RC resonator of Fig. 14.21(b). (Ignore for the moment the additional amplifier drawn with broken lines.) The circuit of Fig. 14.21(b) is a second-order resonator having a pole frequency

$$\omega_0 = 1/\sqrt{LC_6} = 1/\sqrt{C_4C_6R_1R_3R_5/R_2} \tag{14.50}$$

where we have used the expression for L given in Eq. (14.49). The pole Q factor can be obtained using the expression in Eq. (14.40) with $C = C_6$ and $R = R_6$; thus,

$$Q = \omega_0 \, C_6 \, R_6 \tag{14.51}$$

Replacing ω_0 by the expression in Eq. (14.50) gives

$$Q = \omega_0 C_6 R_6 = R_6 \sqrt{\frac{C_6}{C_4} \frac{R_2}{R_1 R_3 R_5}} \tag{14.52}$$

Usually one selects $C_4 = C_6 = C$ and $R_1 = R_2 = R_3 = R_5 = R$, which results in

$$\omega_0 = 1/CR \tag{14.53}$$

$$Q = R_6/R \tag{14.54}$$

Thus, if we select a practically convenient value for C, we can use Eq. (14.53) to determine the value of R to realize a given ω_0, and then use Eq. (14.54) to determine the value of R_6 to realize a given Q.

14.5.3 Realization of the Various Filter Types

The op amp–RC resonator of Fig. 14.21(b) can be used to generate circuit realizations for the various second-order filter functions by following the approach described in Section 14.4 in connection with the LCR resonator. Thus to obtain a bandpass function, we disconnect node z from ground and connect it to the signal source V_i. A high-pass function is obtained by injecting V_i to node y. To realize a low-pass function using the LCR resonator, the inductor terminal x is disconnected from ground and connected to V_i. The corresponding node in the active resonator is the node at which R_5 is connected to ground,[6] labeled as node x in Fig. 14.21(b). A regular notch function ($\omega_n = \omega_0$) is obtained by feeding V_i to nodes x and y. In all cases the output can be taken as the voltage across the resonance circuit, V_r. However, this is not a convenient node to use as the filter output terminal because connecting a load there would change the filter characteristics. The problem can be solved easily by utilizing a buffer amplifier. This is the amplifier of gain K, drawn with broken lines in Fig. 14.21(b). Figure 14.21(c) shows how this amplifier can be simply implemented using an op amp connected in the noninverting configuration. Note that not only does the amplifier K buffer the output of the filter, but it also allows the designer to set the filter gain to any desired value by appropriately selecting the value of K.

Figure 14.22 shows the realizations of the bandpass, low-pass, and general notch filters. Notice that the general notch filter in Fig. 14.22(c) is obtained by feeding the input signal through a portion C_{61} of the capacitance C_6 and a portion R_{51} of the resistance R_5. In order for the poles to remain unchanged, we must ensure that

$$C_{61} + C_{62} = C_6 = C \tag{14.55}$$

and

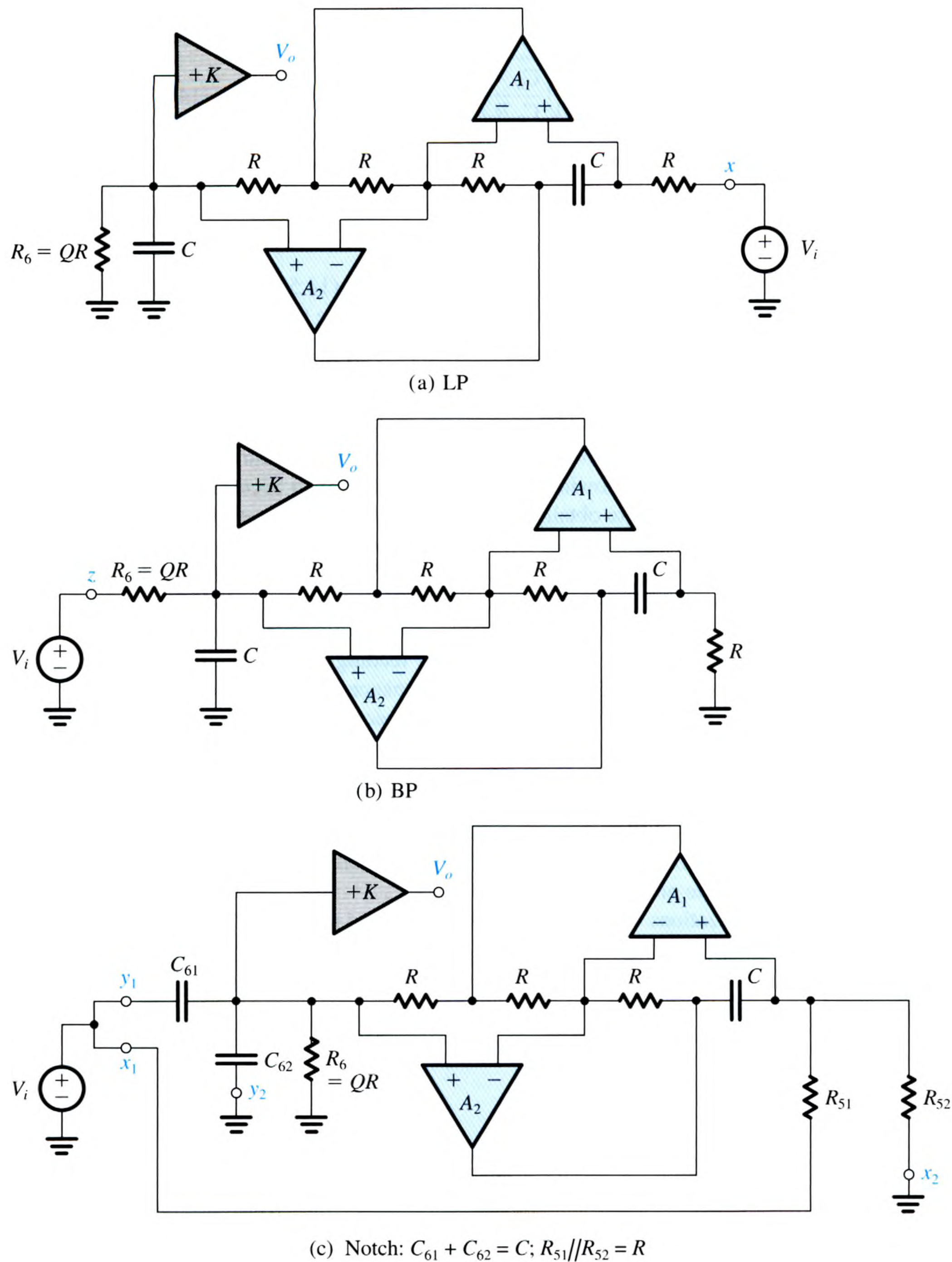
$$R_{51} \| R_{52} = R_5 = R \tag{14.56}$$

The values of C_{61} and R_{51} determine the notch frequency ω_n according to

$$\omega_n = \frac{1}{\sqrt{C_{61} C_4 R_1 R_3 R_{51}/R_2}} \tag{14.57}$$

More details on the design of active filters using the op amp–RC resonator of Fig. 14.21 can be found in Appendix H (on the website).

[6]This point might not be obvious! The reader, however, can show by direct analysis that when V_i is fed to this node, the function V_r/V_i is indeed low pass.



(c) Notch: $C_{61} + C_{62} = C$; $R_{51}//R_{52} = R$

Figure 14.22 Realizations for some second-order filter functions using the op amp–RC resonator of Fig. 14.21(b): **(a)** LP, **(b)** BP, **(c)** general notch. The circuits are based on the resonator circuit in Fig. 14.21.

Example 14.5

Show that the circuit in Fig. 14.22(c) can be used to realize a low-pass notch filter ($\omega_n > \omega_0$) by selecting $R_{52} = \infty$. Design the circuit to obtain $\omega_0 = 10^5$ rad/s, $\omega_n = 1.2 \times 10^5$ rad/s, $Q = 5$, and a dc gain of 5 V/V.

Solution

The notch frequency is given by Eq. (14.57):

$$\omega_n = 1/\sqrt{C_{61} C_4 R_1 R_3 R_{51} / R_2}$$

Selecting $R_{52} = \infty$ results from Eq. (14.56) in $R_{51} = R_5$. Now, using the usual design in which $R_1 = R_2 = R_3 = R_5 = R$ and $C_4 = C_6 = C$ gives

$$\omega_n = \frac{1}{\sqrt{C_{61} C R}}$$

Since ω_0 is given by

$$\omega_0 = \frac{1}{CR}$$

then

$$\frac{\omega_n}{\omega_0} = \sqrt{\frac{C}{C_{61}}} \tag{14.58}$$

Using Eq. (14.55),

$$\frac{\omega_n}{\omega_0} = \sqrt{\frac{C_{61} + C_{62}}{C_{61}}} = \sqrt{1 + \frac{C_{62}}{C_{61}}}$$

Thus, $\omega_n > \omega_0$, and the circuit realizes a low-pass notch (LPN) filter.

For the given numerical values, selecting $C = 1$ nF gives

$$R = \frac{1}{\omega_0 C} = \frac{1}{10^5 \times 10^{-9}} = 10^4\ \Omega = 10\ \text{k}\Omega$$

Since $Q = \omega_0 C R_6$, substituting $Q = 5$ and $C = 1$ nF gives

$$R_6 = \frac{5}{10^5 \times 1 \times 10^{-9}} = 50\ \text{k}\Omega$$

Substituting the given values of ω_0 and ω_n and the value of C into Eq. (14.58) gives

$$\frac{1.2 \times 10^5}{10^5} = \sqrt{\frac{C}{C_{61}}}$$

$$\Rightarrow C_{61} = \frac{C}{1.44} = \frac{1 \times 10^{-9}}{1.44} = 0.694\ \text{nF}$$

Example 14.5 *continued*

Thus,

$$C_{62} = C - C_{61} = 1.0 - 0.694 = 0.306 \text{ nF}$$

Finally, to provide a dc gain of 5, we find the required gain K of the buffer amplifier from

$$\text{dc gain} = K\left(\frac{\omega_n}{\omega_0}\right)^2$$

$$K = \frac{5}{1.44} = 3.47 \text{ V/V}$$

Thus, its two associated resistances can be selected as

$$r_1 = 100 \text{ k}\Omega, \quad r_2 = 247 \text{ k}\Omega$$

EXERCISES

D14.18 Use the Antoniou inductance simulation circuit of Fig. 14.20(a) to realize an inductance of 100 mH.

Ans. Selecting $C = 1$ nF, we obtain $R_1 = R_2 = R_3 = R_5 = 10 \text{ k}\Omega$

D14.19 Use the circuit of Fig. 14.22(b) to design a second-order bandpass filter with a center frequency of 10 kHz, a 3-dB bandwidth of 500 Hz, and a center-frequency gain of 10. Use $C =$ 1.2 nF.

Ans. $R_1 = R_2 = R_3 = R_5 = 13.26$ kΩ; $R_6 = 265$ kΩ; $C_4 = C_6 = 1.2$ nF; $K = 10$, $r_1 = 10$ kΩ, $r_2 = 90$ kΩ

14.6 Second-Order Active Filters Based on the Two-Integrator Loop

In this section, we study another family of op amp–RC circuits that realize second-order filter functions. The circuits are based on the use of two integrators connected in cascade in an overall feedback loop and are thus known as two-integrator-loop circuits.

14.6.1 Derivation of the Two-Integrator-Loop Biquad

To derive the two-integrator-loop biquadratic circuit, or **biquad** as it is commonly known,[7] consider the second-order high-pass transfer function

$$\frac{V_{hp}}{V_i} = \frac{Ks^2}{s^2 + s(\omega_0/Q) + \omega_0^2} \tag{14.59}$$

where K is the high-frequency gain. Cross-multiplying Eq. (14.59) and dividing both sides of the resulting equation by s^2 (to get all the terms involving s in the form $1/s$, which is the transfer function of an integrator) gives

$$V_{hp} + \frac{1}{Q}\left(\frac{\omega_0}{s}V_{hp}\right) + \left(\frac{\omega_0^2}{s^2}V_{hp}\right) = KV_i \tag{14.60}$$

In this equation we observe that the signal $(\omega_0/s)V_{hp}$ can be obtained by passing V_{hp} through an integrator with a time constant equal to $1/\omega_0$. Furthermore, passing the resulting signal through another identical integrator results in the third signal involving V_{hp} in Eq. (14.60)—namely, $(\omega_0^2/s^2)V_{hp}$. Figure 14.23(a) shows a block diagram for such a two-integrator arrangement. Note that in anticipation of the use of the inverting op-amp

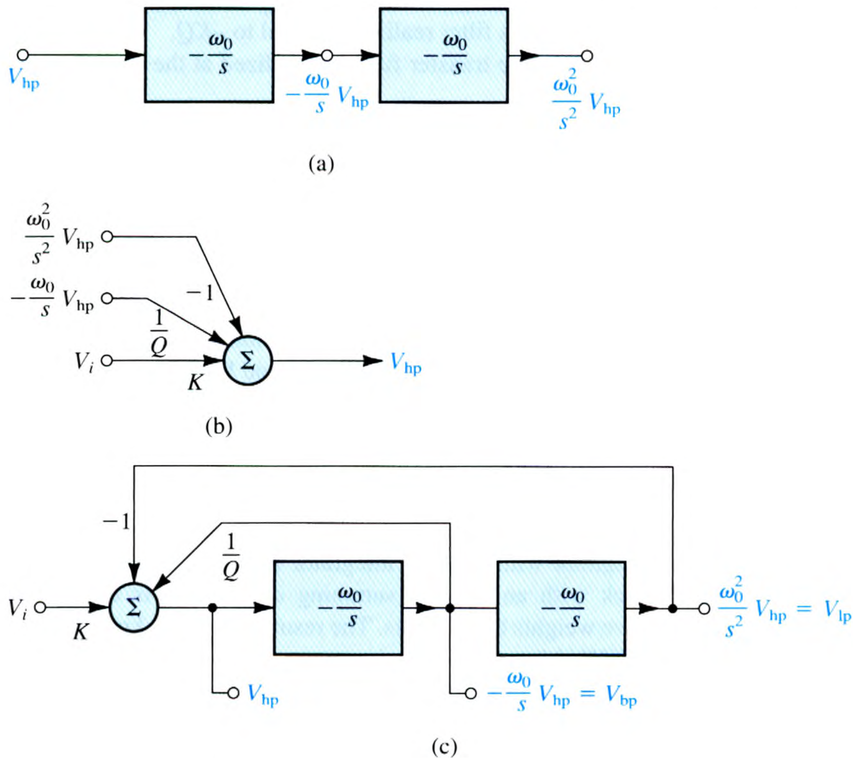


Figure 14.23 Derivation of a block diagram realization of the two-integrator-loop biquad.

[7]The name *biquad* stems from the fact that this circuit in its most general form is capable of realizing a biquadratic transfer function, that is, one that is the ratio of two quadratic polynomials.

Miller integrator circuit (Section 2.5.2) to implement each integrator, the integrator blocks in Fig. 14.23(a) have been assigned negative signs.

The problem still remains, however, of how to form V_{hp}, the input signal feeding the two cascaded integrators. Toward that end, we rearrange Eq. (14.60), expressing V_{hp} in terms of its single- and double-integrated versions and of V_i as

$$V_{hp} = KV_i - \frac{1}{Q}\frac{\omega_0}{s}V_{hp} - \frac{\omega_0^2}{s^2}V_{hp} \tag{14.61}$$

which suggests that V_{hp} can be obtained by using the weighted summer of Fig. 14.23(b). Now it should be easy to see that a complete block diagram realization can be obtained by combining the integrator blocks of Fig. 14.23(a) with the summer block of Fig. 14.23(b), as shown in Fig. 14.23(c).

In the realization of Fig. 14.23(c), V_{hp}, obtained at the output of the summer, realizes the high-pass transfer function $T_{hp} \equiv V_{hp}/V_i$ of Eq. (14.59). The signal at the output of the first integrator is $-(\omega_0/s)V_{hp}$, which is a bandpass function,

$$\frac{(-\omega_0/s)V_{hp}}{V_i} = -\frac{K\omega_0 s}{s^2 + s(\omega_0/Q) + \omega_0^2} = T_{bp}(s) \tag{14.62}$$

Therefore the signal at the output of the first integrator is labeled V_{bp}. Note from Eq. (14.62) that the center-frequency gain of the bandpass filter realized is equal to $-KQ$.

In a similar fashion, we can show that the transfer function realized at the output of the second integrator is the low-pass function,

$$\frac{(\omega_0^2/s^2)V_{hp}}{V_i} = \frac{K\omega_0^2}{s^2 + s(\omega_0/Q) + \omega_0^2} = T_{lp}(s) \tag{14.63}$$

Thus the output of the second integrator is labeled V_{lp}. Note that the dc gain of the low-pass filter realized is equal to K.

We conclude that the two-integrator-loop biquad shown in block diagram form in Fig. 14.23(c) realizes the three basic second-order filtering functions, LP, BP, and HP, *simultaneously*. This versatility has made the circuit very popular and has given it the name *universal active filter*.

14.6.2 Circuit Implementation

To obtain an op-amp circuit implementation of the two-integrator-loop biquad of Fig. 14.23(c), we replace each integrator with a Miller integrator circuit having $CR = 1/\omega_0$, and we replace the summer block with an op-amp summing circuit that is capable of assigning both positive and negative weights to its inputs. The resulting circuit, known as the Kerwin–Huelsman–Newcomb or **KHN biquad**, after its inventors, is shown in Fig. 14.24(a). Given required values for ω_0, Q, and K, the design of the circuit is straightforward: We select suitably practical values for the components C and R of the integrators so that $CR = 1/\omega_0$. To determine the required values of the resistors associated with the summer, we first use *superposition* to express the output of the summer V_{hp} in terms of its inputs, V_i, V_{bp}, and V_{lp} as

$$V_{hp} = V_i \frac{R_3}{R_2 + R_3}\left(1 + \frac{R_f}{R_1}\right) + V_{bp}\frac{R_2}{R_2 + R_3}\left(1 + \frac{R_f}{R_1}\right) - V_{lp}\frac{R_f}{R_1}$$

Substituting $V_{bp} = -(\omega_0/s)V_{hp}$ and $V_{lp} = (\omega_0{}^2/s^2)V_{hp}$ gives

$$V_{hp} = \frac{R_3}{R_2+R_3}\left(1+\frac{R_f}{R_1}\right)V_i + \frac{R_2}{R_2+R_3}\left(1+\frac{R_f}{R_1}\right)\left(-\frac{\omega_0}{s}V_{hp}\right) - \frac{R_f}{R_1}\left(\frac{\omega_0^2}{s^2}V_{hp}\right) \tag{14.64}$$

Equating the last right-hand-side terms of Eqs. (14.64) and (14.61) gives

$$R_f/R_1 = 1 \tag{14.65}$$

which implies that we can select arbitrary but practically convenient equal values for R_1 and R_f. Then, equating the second-to-last terms on the right-hand side of Eqs. (14.64) and (14.61) and setting $R_1 = R_f$ yields the ratio R_3/R_2 required to realize a given Q as

$$R_3/R_2 = 2Q - 1 \tag{14.66}$$

Thus an arbitrary but convenient value can be selected for either R_2 or R_3, and the value of the other resistance can be determined using Eq. (14.66). Finally, equating the coefficients of V_i in Eqs. (14.64) and (14.61) and substituting $R_f = R_1$ and for R_3/R_2 from Eq. (14.65) results in

$$K = 2 - (1/Q) \tag{14.67}$$

Thus the gain parameter K is fixed to this value.

The KHN biquad can be used to realize arbitrary transmission zeros including those for notch filters by summing weighted versions of the three outputs, LP, BP, and HP. Such an op-amp summer is shown in Fig. 14.24(b). (See Problem 14.61.)

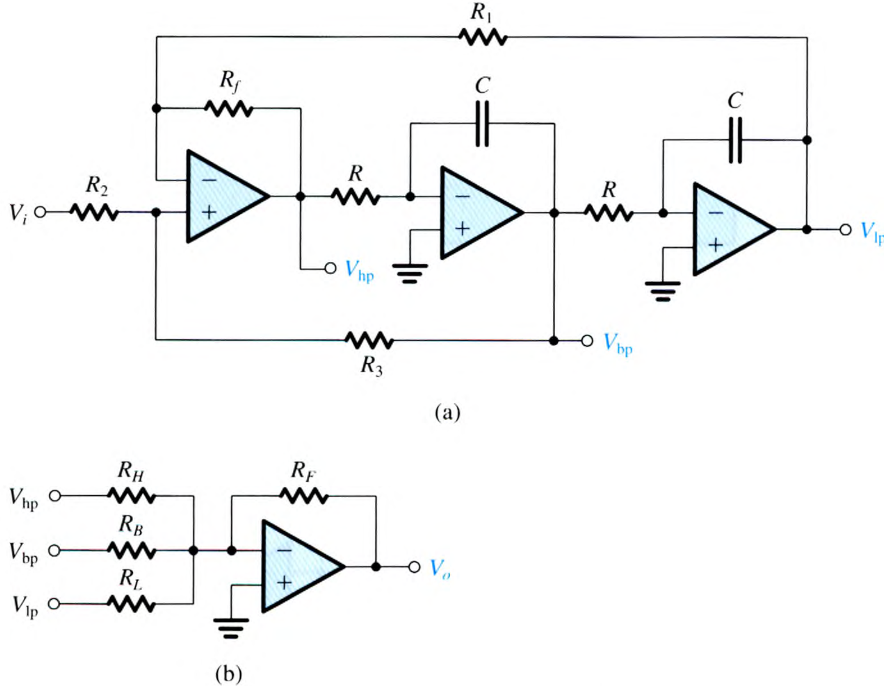


Figure 14.24 **(a)** The KHN biquad circuit, obtained as a direct implementation of the block diagram of Fig. 14.23(c). The three basic filtering functions, HP, BP, and LP, are simultaneously realized. **(b)** To obtain notch and other functions, the three outputs are summed with appropriate weights using this op-amp summer.

EXERCISE

D 14.20 Design the KHN circuit to realize a high-pass function with $f_0 = 10$ kHz and $Q = 2$. Choose $C = 1$ nF. What is the value of high-frequency gain obtained? What is the center-frequency gain of the bandpass function that is simultaneously available at the output of the first integrator?
Ans. $R = 15.9$ kΩ; $R_1 = R_f = R_2 = 10$ kΩ (arbitrary); $R_3 = 30$ kΩ; 1.5 V/V; 3 V/V

14.6.3 An Alternative Two-Integrator-Loop Biquad Circuit

There is another way to develop a two-integrator-loop biquad circuit with all three op amps used in a single-ended mode (i.e., with noninverting input terminal grounded). Rather than using the input summer to add signals with positive and negative coefficients [Fig. 14.23(c)], we can introduce an additional inverter, as shown in Fig. 14.25(a). Now all the coefficients of the summer have the same sign, and we can dispense with the summing amplifier altogether and perform the summation at the virtual-ground input of the first integrator as shown in Fig. 14.25(b). Observe that the summing weights of 1, $1/Q$, and K are realized by using resistances of R, QR, and R/K, respectively. The resulting circuit is shown in Fig. 14.25(b), from which we observe that the high-pass function is no longer available! This is the price

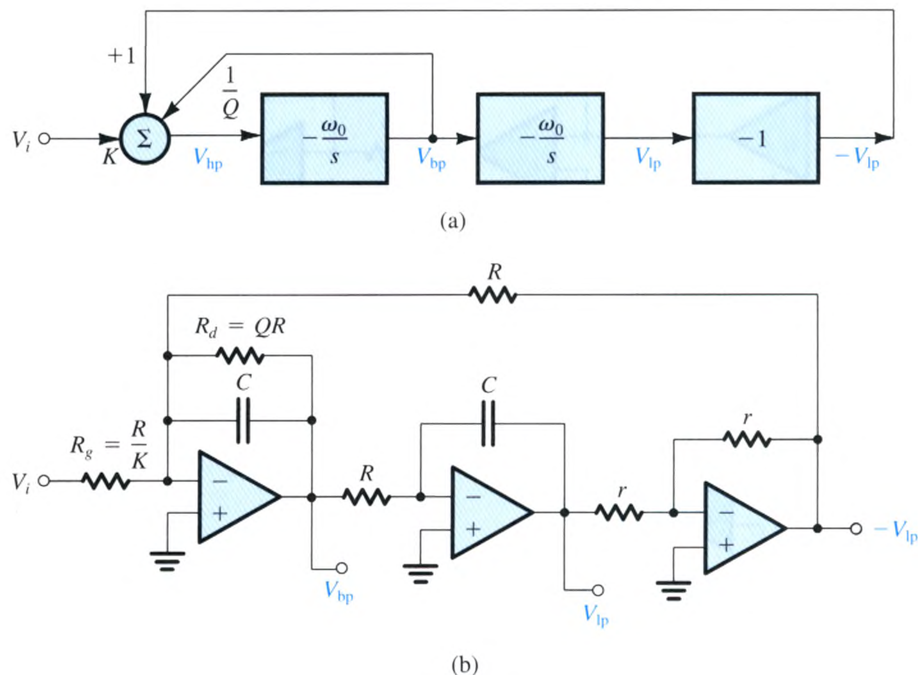


Figure 14.25 **(a)** Derivation of an alternative two-integrator-loop biquad in which all op amps are used in a single-ended fashion. **(b)** The resulting circuit, known as the Tow–Thomas biquad.

paid for obtaining a circuit that utilizes all op amps in a single-ended mode. The circuit of Fig. 14.25(b) is known as the **Tow–Thomas biquad**, after its originators.

Here the bandpass tranfer function is given by Eq. (14.62),

$$T_{bp}(s) = \frac{V_{bp}}{V_i} = -\frac{K\omega_0 s}{s^2 + s(\omega_0/Q) + \omega_0^2} \tag{14.68}$$

and the low-pass function is given by Eq. (14.63)

$$T_{lp}(s) = \frac{V_{lp}}{V_i} = \frac{K\omega_0^2}{s^2 + s(\omega_0/Q) + \omega_0^2} \tag{14.69}$$

where

$$\omega_0 = 1/CR$$

Rather than using a fourth op amp to realize the finite transmission zeros required for the notch and other functions, as was done with the KHN biquad, an economical *feedforward* scheme can be employed with the Tow–Thomas circuit. Specifically, the virtual ground available at the input of each of the three op amps in the Tow–Thomas circuit permits the input signal to be fed to all three op amps, as shown in Fig. 14.26. If V_o is taken at the output of the damped integrator, straightforward analysis yields the filter transfer function

$$\frac{V_o}{V_i} = -\frac{s^2\left(\dfrac{C_1}{C}\right) + s\dfrac{1}{C}\left(\dfrac{1}{R_1} - \dfrac{r}{RR_3}\right) + \dfrac{1}{C^2RR_2}}{s^2 + s\dfrac{1}{QCR} + \dfrac{1}{C^2R^2}} \tag{14.70}$$

which as an example can be used to obtain a notch function by selecting $R_1 = R_3 = \infty$ (i.e., omitting R_1 and R_3 altogether). Table H.2 in Appendix H on the website provides design equations for a variety of filter functions realized using the circuit in Fig. 14.26. See also Problems 14.63–14.66.

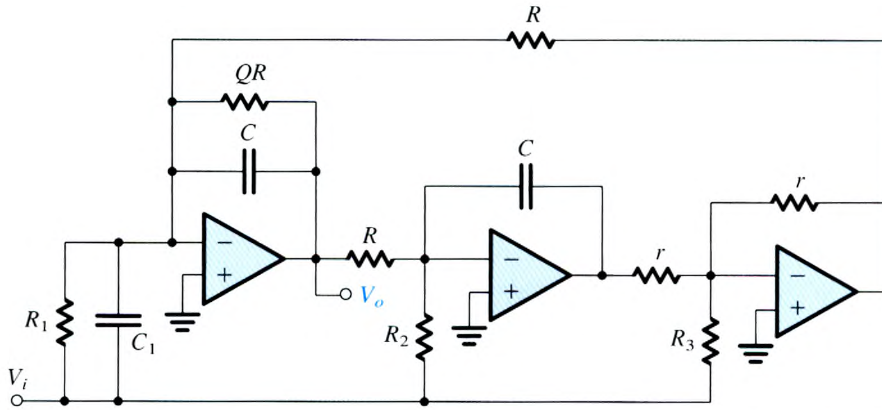


Figure 14.26 The Tow–Thomas biquad with feedforward. The transfer function of Eq. (14.70) is realized by feeding the input signal through appropriate components to the inputs of the three op amps. This circuit can realize all special second-order functions. The design equations are given in Table H.2 in Appendix H on the website.

14.6.4 Final Remarks

Two-integrator-loop biquads are extremely versatile and easy to design. However, their performance is adversely affected by the finite bandwidth of the op amps. Special techniques exist for compensating the circuit for such effects (see the SPICE simulation example on the website, and Sedra and Brackett, 1978).

Example 14.6

Show that the Tow–Thomas biquad circuit in Fig. 14.26 can be used to realize a notch filter by selecting $R_1 = R_3 = \infty$. What is the frequency of the notch? Design the circuit to obtain $\omega_0 = 10^5$ rad/s, $\omega_n = 1.2 \times 10^5$ rad/s, $Q = 5$, and a dc gain of -5 V/V.

Solution

Substituting $R_1 = R_3 = \infty$ into Eq. (14.70) gives us

$$\frac{V_o}{V_i} = -\frac{s^2\left(\frac{C_1}{C}\right) + \frac{1}{C^2RR_2}}{s^2 + s\frac{1}{QCR} + \frac{1}{C^2R^2}} \tag{14.71}$$

which is the transfer function of a notch filter having

$$\omega_n = \sqrt{\frac{1}{C^2RR_2}\frac{C}{C_1}} = 1/\sqrt{(CR)(C_1R_2)} \tag{14.72}$$

$$\omega_0 = 1/CR \tag{14.73}$$

and

$$\text{dc gain} = -\frac{R}{R_2} \tag{14.74}$$

To obtain $\omega_0 = 10^5$ rad/s, $\omega_n = 1.2 \times 10^5$ rad/s, $Q = 5$, and dc gain $= -5$ V/V, we select $C = 1$ nF and obtain the value of R using Eq. (14.73):

$$10^5 = \frac{1}{10^{-9} \times R}$$

$$\Rightarrow R = 10\ \text{k}\Omega$$

We can now find the value of R_2 using Eq. (14.74):

$$-5 = -\frac{10}{R_2}$$

$$\Rightarrow R_2 = 2\ \text{k}\Omega$$

Then, we can determine the value of C_1 using Eq. (14.72):

$$1.2 \times 10^5 = \frac{1}{\sqrt{10^{-9} \times 10^4 \times C_1 \times 2 \times 10^3}}$$

$$\Rightarrow C_1 = 3.472\ \text{nF}$$

Finally, we can find the value of the integrator damping resistance:

$$QR = 5 \times 10 = 50\ \text{k}\Omega$$

EXERCISE

D14.21 Use the Tow–Thomas biquad [Fig. 14.25(b)] to design a second-order bandpass filter with $f_0 =$ 10 kHz, $Q = 20$, and unity center-frequency gain. If $R = 10$ kΩ, give the values of C, R_d, and R_g.
Ans. 1.59 nF; 200 kΩ; 200 kΩ

14.7 Second-Order Active Filters Using a Single Op Amp

The op amp–RC second-order filter circuits studied in the two preceding sections provide good performance, are versatile, and are easy to design and to adjust (tune) after final assembly. Unfortunately, they are not economical in their use of op amps, requiring three or four amplifiers per second-order section. This can be a problem, especially in applications that call for conservation of power-supply current: for instance, in a battery-operated instrument. In this section we study second-order filter circuits that require only one op amp per circuit. These minimal realizations, however, suffer a greater dependence on the limited gain and bandwidth of the op amp than the multiple-op-amp circuits of the preceding sections. Single-amplifier circuits are therefore limited to the less stringent filter specifications—for example, pole Q factors less than about 5.

Although two-port circuits that contain only resistors and capacitors can have complex-conjugate zeros, their poles are restricted to lie on the negative real axis of the s plane, and so their Q factors are less than 0.5. Thus, RC circuits alone cannot be used to implement selective filters, which generally require complex-conjugate poles. However, by placing a two-port RC circuit in the feedback path of an op amp, transfer functions with complex-conjugate poles can be realized. For space limitations, we will not study the theory for generating single-amplifier biquadratic filter circuits (or SABs). Instead, we present single-amplifier circuit realizations for the basic second-order filter functions: bandpass, low-pass, and high-pass.

14.7.1 Bandpass Circuit

Figure 14.27(a) shows a single-amplifier circuit that realizes a second-order bandpass transfer function. We can perform the following inspection to verify that it is, in fact, a bandpass circuit: At dc, C_1 and C_2 behave as open circuits, causing V_o to equal the voltage at the inverting input terminal of the amplifier and thus zero. At $s = \infty$, C_1 and C_2 behave as short circuits, connecting the output terminal to the inverting op-amp input and thus forcing V_o to equal zero. Thus, there is a transmission zero at $s = 0$ and another at $s = \infty$, which is the case for a second-order bandpass.

The transfer function of the circuit in Fig. 14.27(a) can be derived as shown in Fig. 14.27(b). Writing a node equation at X gives

$$\frac{V_o}{R_3} + sC_1\left(V_o + \frac{1}{C_2R_3}V_o\right) + \frac{1}{R_4}\left(V_i + \frac{1}{sC_2R_3}V_o\right) = 0$$

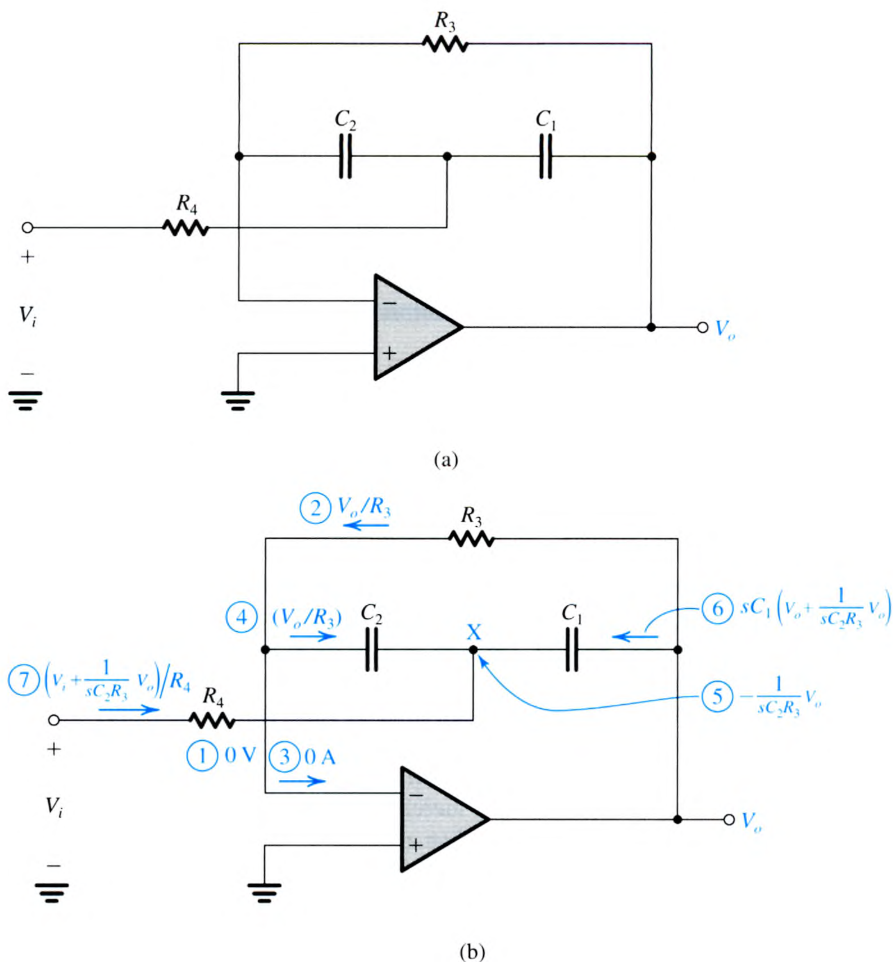


Figure 14.27 **(a)** Circuit for the realization of a second-order bandpass filter. **(b)** Analysis of the circuit in (a) to determine the transfer function $T(s) \equiv V_o/V_i$. The order of the analysis steps is indicated by the circled numbers

Collecting terms, we can find the transfer function V_o/V_i and write it as follows:

$$T(s) \equiv \frac{V_o}{V_i} = -\frac{s\dfrac{1}{C_1R_4}}{s^2 + s\dfrac{1}{R_3}\left(\dfrac{1}{C_1} + \dfrac{1}{C_2}\right) + \dfrac{1}{C_1C_2R_3R_4}} \tag{14.75}$$

If we compare this with the standard form of the second-order bandpass function

$$T(s) = \frac{sK(\omega_0/Q)}{s^2 + s\left(\dfrac{\omega_0}{Q}\right) + \omega_0^2} \tag{14.76}$$

we get

$$\omega_0 = \frac{1}{\sqrt{C_1 C_2 R_3 R_4}} \tag{14.77}$$

$$Q = \frac{1}{\sqrt{C_1 C_2 R_3 R_4}} \Big/ \left[\frac{1}{R_3}\left(\frac{1}{C_1} + \frac{1}{C_2}\right)\right] \tag{14.78}$$

$$K = -\frac{R_3}{R_4} \Big/ \left(1 + \frac{C_1}{C_2}\right) \tag{14.79}$$

Given required values for ω_0 and Q, we can determine the component values of the circuit in Fig. 14.27(a) as follows: Usually we select

$$C_1 = C_2 = C \tag{14.80}$$

$$R_3 = R \tag{14.81}$$

$$R_4 = R/m \tag{14.82}$$

where m can be determined by substituting these values into Eq. (14.78):

$$m = 4Q^2 \tag{14.83}$$

Now, using Eq. (14.77), we can find the required value of the time constant CR:

$$CR = \frac{2Q}{\omega_0} \tag{14.84}$$

To summarize, we determine the required time-constant CR using Eq. (14.84). Then, selecting a practically convenient value for C, we determine $C_1 = C_2 = C$ and the required value for R. Finally, $R_3 = R$ and $R_4 = R/4Q^2$. The center-frequency gain can be determined from Eq. (14.79):

$$K = -2Q^2 \tag{14.85}$$

EXERCISE

D14.22 Design the circuit in Fig. 14.27(a) to realize a bandpass filter with $\omega_0 = 10^5$ rad/s and $Q = 5$. What center-frequency gain do we obtain?
Ans. $C_1 = C_2 = 1$ nF; $R_3 = 100$ kΩ; $R_4 = 1$ kΩ; $K = -50$ V/V

14.7.2 High-Pass Circuit

Figure 14.28(a) shows a single-amplifier circuit that realizes the second-order high-pass filter function

$$T(s) = \frac{K s^2}{s^2 + s\left(\frac{\omega_0}{Q}\right) + \omega_0^2} \tag{14.86}$$

Notice that here the op amp is connected as a unity-gain voltage follower and can be replaced with any other unity-gain follower circuit such as a source follower, if desired.

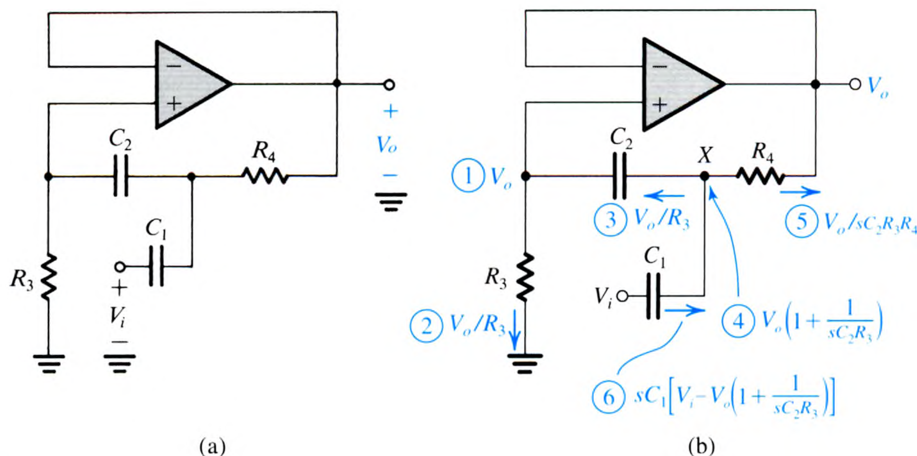


Figure 14.28 **(a)** Circuit for realizing the second-order high-pass filter function. **(b)** Analysis of the circuit in (a) to determine its transfer function $T(s) = V_o/V_i$. The circled numbers indicate the order of the analysis steps.

Take a moment to recognize that at $s = 0$ (dc), C_1 and C_2 present two *independent* reasons for making $V_o = 0$. Thus, the circuit has two transmission zeros at $s = 0$, as a second-order high-pass circuit must.

The transfer function $T(s) = V_o/V_i$ can be derived as illustrated in Fig. 14.28(b). Writing a node equation at X given

$$T(s) \equiv \frac{V_o}{V_i} = \frac{s^2}{s^2 + s\frac{1}{R_3}(\frac{1}{C_1} + \frac{1}{C_2}) + \frac{1}{C_1 C_2 R_3 R_4}} \tag{14.87}$$

We observe that the denominator polynomial is identical to that in Eq. (14.75). Thus, the expressions for ω_0 and Q are the same as those in Eqs. (14.77) and (14.78), respectively. As well, given desired values for ω_0 and Q, the design of the circuit in Fig. 14.28(a) will be identical to the one we developed in Section 14.7.1 for the circuit in Fig. 14.27(a):

$$C_1 = C_2 = C \tag{14.88}$$

$$R_3 = R \tag{14.89}$$

$$R_4 = R/4\,Q^2 \tag{14.90}$$

$$CR = 2\,Q/\omega_0 \tag{14.91}$$

where C is chosen to be a practically convenient value. Comparing Eqs. (14.87) and (14.86), we see that the high-frequency gain realized is

$$K = 1 \tag{14.92}$$

This result should not be surprising: From the circuit in Fig. 14.28(a), we can see that as $s \to \infty$, C_1 and C_2 behave as short circuits, making the voltage at the noninverting input of the op amp equal to V_i, and so $V_o = V_i$ and $K = 1$.

Finally, note that the fact that the circuits in Fig. 14.27(a) and Fig. 14.28(a) have the same poles is not an accident. These two circuits are related to each other through the **complementary transformation**. More on this can be found on the companion website.

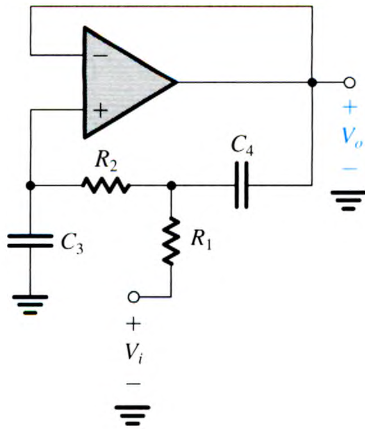


Figure 14.29 A single-amplifier circuit that realizes the second-order low-pass filter function.

14.7.3 Low-Pass Circuit

Figure 14.29 shows a single-amplifier circuit that realizes the second-order low-pass filter function

$$T(s) = \frac{K\,\omega_0^2}{s^2 + s\,\dfrac{\omega_0}{Q} + \omega_0^2} \tag{14.93}$$

We can verify that this is a low-pass circuit by observing that at $s = \infty$, there are two independent reasons for V_o to equal zero: C_3 acting as a short circuit and, independently, C_4 acting as a short circuit. Thus, the circuit has two transmission zeros at $s = \infty$, as a second-order low-pass filter must.

We can analyze the circuit in Fig. 14.29 to determine its transfer function $T(s)$ using a method similar to the one we used for the high-pass circuit in Fig. 14.28(b). The result is

$$T(s) \equiv \frac{V_o}{V_i} = \frac{1/C_3C_4\,R_1R_2}{s^2 + s\,\frac{1}{C_4}\left(\dfrac{1}{R_1} + \dfrac{1}{R_2}\right) + \dfrac{1}{C_3C_4R_1R_2}} \tag{14.94}$$

Comparing Eqs. (14.94) and (14.93) gives us

$$\omega_0 = 1/\sqrt{C_3C_4R_1R_2} \tag{14.95}$$

$$Q = \frac{1}{\sqrt{C_3C_4R_1R_2}}\Bigg/\left[\frac{1}{C_4}\left(\frac{1}{R_1} + \frac{1}{R_2}\right)\right] \tag{14.96}$$

$$\text{dc gain} = 1 \tag{14.97}$$

Typically, this circuit is designed as follows:

$$R_1 = R_2 = R \tag{14.98}$$

$$C_4 = C \tag{14.99}$$

$$C_3 = C/4Q^2 \tag{14.100}$$

$$CR = \frac{2Q}{\omega_0} \tag{14.101}$$

where either R or C is chosen to yield practically convenient component values.

EXERCISE

D14.23 Design the circuit in Fig. 14.29 to realize a second-order low-pass filter with $\omega_0 = 10^5$ rad/s and $Q = 1/\sqrt{2}$. (Note that this is a Butterworth filter of order 2.) What is the dc gain realized?
Ans. $R_1 = R_2 = 14.14$ kΩ; $C_3 = 0.5$ nF; $C_4 = 1$ nF; $K = 1$

14.8 Switched-Capacitor Filters

In this section we study an important approach to the design of analog filters for IC implementation. Switched-capacitor filters, which require only small capacitors, analog switches, and op amps that need to drive only small capacitive loads, are ideally suited for implementation in CMOS. Currently, the switched-capacitor approach is the preferred method for the design of integrated-circuit filters in the frequency range up to hundreds of kHz.

14.8.1 The Basic Principle

The switched-capacitor filter technique is based on the realization that a capacitor switched between two circuit nodes at a sufficiently high rate is equivalent to a resistor connecting these two nodes. To be specific, consider the active-RC integrator of Fig. 14.30(a). This is the familiar Miller integrator, which we used in the two-integrator-loop biquad in Section 14.6. In Fig. 14.30(b) we have replaced the input resistor R_1 by a grounded capacitor C_1 together with two MOS transistors acting as switches. In some circuits, more elaborate switch configurations are used, but such details are beyond our present need.

The two MOS switches in Fig. 14.30(b) are driven by a *nonoverlapping* two-phase clock. Figure 14.30(c) shows the clock waveforms. For now, let's assume that the clock frequency $f_c (f_c = 1/T_c)$ is much higher than the frequency of the input signal v_i. Thus the variations in the input signal are negligibly small during clock phase ϕ_1, when C_1 is connected across the input signal source v_i. It follows that during ϕ_1, capacitor C_1 charges up to the voltage v_i,

$$q_{C1} = C_1 v_i$$

Then, during clock phase ϕ_2, capacitor C_1 is connected to the virtual-ground input of the op amp, as indicated in Fig. 14.30(d). Capacitor C_1 is thus forced to discharge, and its previous charge q_{C1} is transferred to C_2, in the direction indicated in Fig. 14.30(d).

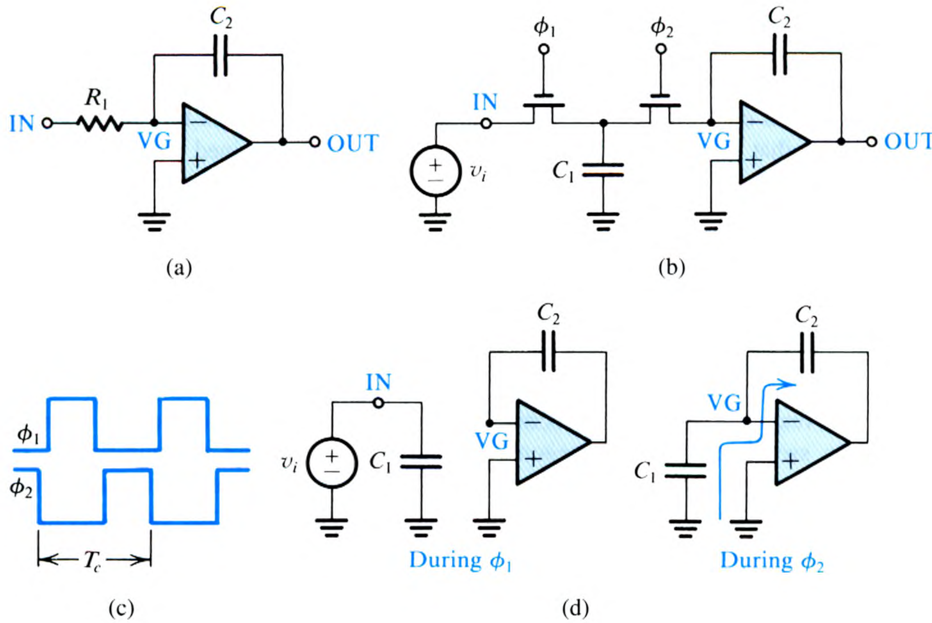


Figure 14.30 Basic principle of the switched-capacitor filter technique. **(a)** Active-RC integrator. **(b)** Switched-capacitor integrator. **(c)** Two-phase clock (nonoverlapping). **(d)** During ϕ_1, C_1 charges up to the current value of v_i and then, during ϕ_2, discharges into C_2.

From the description above we see that during each clock period T_c an amount of charge $q_{C1} = C_1 v_i$ is extracted from the input source and supplied to the integrator capacitor C_2. Thus the average current flowing between the input node (IN) and the virtual-ground node (VG) is

$$i_{\text{av}} = \frac{C_1 v_i}{T_c}$$

If T_c is short enough, we can think of this process as almost continuous and thus can define an equivalent resistance R_{eq} that is in effect present between nodes IN and VG:

$$R_{\text{eq}} \equiv v_i / i_{\text{av}}$$

Thus,

$$R_{\text{eq}} = T_c / C_1 \tag{14.102}$$

Using R_{eq} we obtain an equivalent time constant for the integrator:

$$\text{Time constant} = C_2 R_{\text{eq}} = T_c \frac{C_2}{C_1} \tag{14.103}$$

Thus the time constant of the integrator, which determines the frequency response of the filter, is established by the clock period T_c and the capacitor ratio C_2/C_1. Both these parameters can be well controlled in an IC fabrication process. Specifically, note the dependence on capacitor ratios rather than on absolute values of capacitors. The accuracy of capacitor ratios in CMOS technology can be controlled to within 0.1%.

Another point worth observing is that with a reasonable clocking frequency (such as 100 kHz) and not-too-large capacitor ratios (say, 10), we can obtain reasonably large time constants (such as 10^{-4} s) suitable for audio applications. Since capacitors typically occupy relatively large areas on the IC chip, we try to minimize their values. In this context, it is important to note that the ratio accuracies quoted earlier can be obtained with the smaller capacitor value as low as 0.1 pF.

14.8.2 Switched-Capacitor Integrator

The switched-capacitor (SC) circuit in Fig. 14.30(b) realizes an inverting integrator [note the direction of charge flow through C_2 in Fig. 14.30(d)]. As we saw in Section 14.6, a two-integrator-loop active filter is composed of one inverting and one noninverting integrator.[8] To realize a switched-capacitor biquad filter, we therefore need a pair of complementary switched-capacitor integrators. Figure 14.31(a) shows a noninverting, or positive, integrator circuit. We encourage you to follow the operation of this circuit during the two clock phases to see that it operates in much the same way as the basic circuit of Fig. 14.30(b), except for a sign reversal.

In addition to realizing a noninverting integrator function, the circuit in Fig. 14.31(a) is insensitive to stray capacitances. We will not explore this point any further here, but you may consult Schaumann, Ghausi, and Laker (1990) for more information. By reversing the clock phases on two of the switches, we get the circuit in Fig. 14.31(b). This circuit realizes the inverting integrator function, like the circuit of Fig. 14.30(b), but is insensitive to stray capacitances [which the original circuit of Fig. 14.30(b) is not]. The complementary integrators of Fig. 14.31 have become the standard building blocks in the design of switched-capacitor filters.

14.8.3 Switched-Capacitor Biquad Filter

Let us now consider the realization of a complete biquad circuit. Figure 14.32(a) shows the active-RC, two-integrator-loop circuit studied earlier. By considering the cascade of integrator 2 and the inverter as a positive integrator, and then simply replacing each resistor by its switched-capacitor equivalent, we obtain the circuit in Fig. 14.32(b). Ignore the damping around the first integrator (i.e., the switched capacitor C_5) for the time being and note that the feedback loop indeed consists of one inverting and one noninverting integrator. Then note the phasing of the switched capacitor used for damping. Reversing the phases here would convert the feedback to positive and move the poles to the right half of the s plane. On the other hand, the phasing of the feed-in switched capacitor (C_6) is not that important; a reversal of phases would result only in an inversion in the sign of the function realized.

Having identified the correspondences between the active-RC biquad and the switched-capacitor biquad, we can now derive design equations. Analysis of the circuit in Fig. 14.32(a) yields

$$\omega_0 = \frac{1}{\sqrt{C_1 C_2 R_3 R_4}} \tag{14.104}$$

[8] In the two-integrator loop of Fig. 14.25(b), the noninverting integrator is realized by the cascade of a Miller integrator and an inverting amplifier.

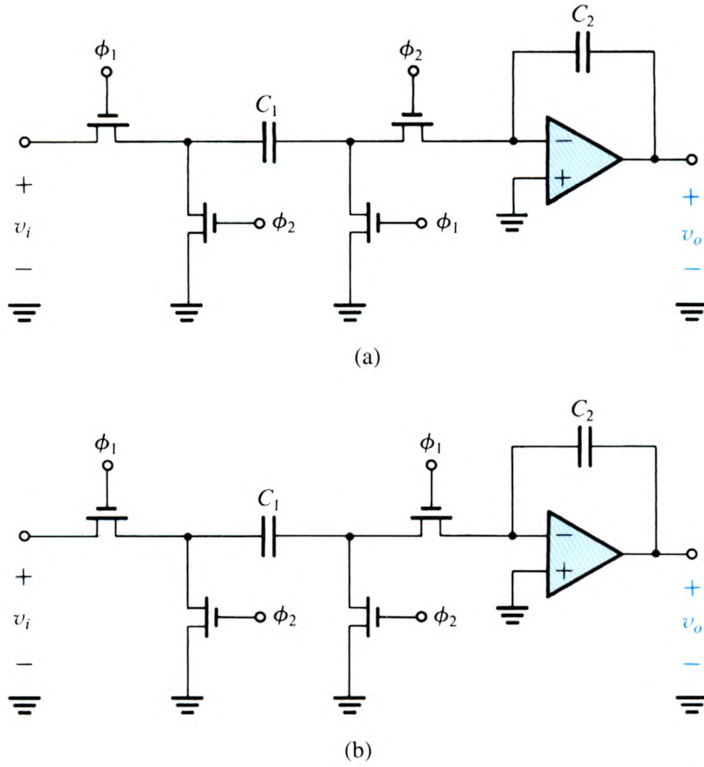


Figure 14.31 A pair of complementary stray-insensitive, switched-capacitor integrators. **(a)** Noninverting switched-capacitor integrator. **(b)** Inverting switched-capacitor integrator.

Replacing R_3 and R_4 with their switched-capacitor equivalent values, that is,

$$R_3 = T_c/C_3 \quad \text{and} \quad R_4 = T_c/C_4$$

gives ω_0 of the switched-capacitor biquad as

$$\omega_0 = \frac{1}{T_c}\sqrt{\frac{C_3}{C_2}\frac{C_4}{C_1}} \tag{14.105}$$

Typically we select the time constants of the two integrators to be equal; that is,

$$\frac{T_c}{C_3}C_2 = \frac{T_c}{C_4}C_1 \tag{14.106}$$

If, further, we select the two integrating capacitors C_1 and C_2 to be equal,

$$C_1 = C_2 = C \tag{14.107}$$

then, using Eqs. (14.106) and (14.105), we obtain

$$C_3 = C_4 = (\omega_0 T_c)C \tag{14.108}$$

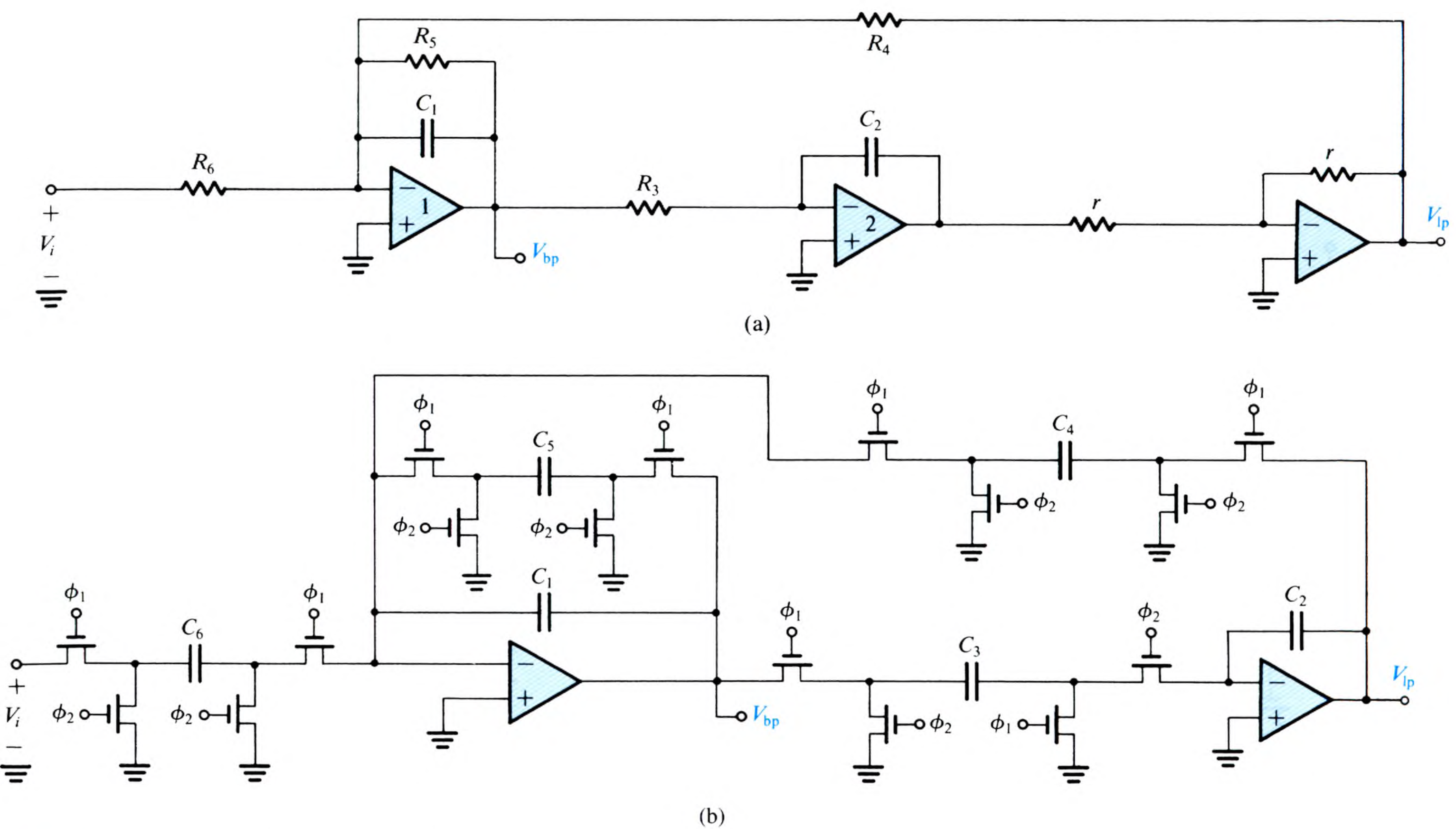


Figure 14.32 **(a)** A two-integrator-loop, active-RC biquad and **(b)** its switched-capacitor counterpart.

For the case of equal time constants, the Q factor of the circuit in Fig. 14.32(a) is given by R_5/R_4. Thus the Q factor of the corresponding switched-capacitor circuit in Fig. 14.32(b) is given by

$$Q = \frac{T_c/C_5}{T_c/C_4} = \frac{C_4}{C_5} \tag{14.109}$$

Thus C_5 should be selected from

$$C_5 = \frac{C_4}{Q} = \left(\frac{\omega_0 T_c}{Q}\right)C \tag{14.110}$$

Finally, the center-frequency gain of the bandpass function is given by

$$\text{Center-frequency gain} = \frac{C_6}{C_5} = \left(\frac{Q}{\omega_0 T_c}\right)\left(\frac{C_6}{C}\right) \tag{14.111}$$

EXERCISE

D14.24 Use $C_1 = C_2 = 20$ pF and design the circuit in Fig. 14.32(b) to realize a bandpass function with $f_0 = 10$ kHz, $Q = 20$, and unity center-frequency gain. Use a clock frequency $f_c = 200$ kHz. Find the values of C_3, C_4, C_5, and C_6.
Ans. 6.283 pF; 6.283 pF; 0.314 pF; 0.314 pF

14.8.4 Final Remarks

We have attempted to provide only an introduction to switched-capacitor filters. We have made many simplifying assumptions, the most important being the switched-capacitor–resistor equivalence (Eq. 14.102). This equivalence is correct only at $f_c = \infty$ and is approximately correct for $f_c \gg f_0$. Switched-capacitor filters are, in fact, discrete-time circuits whose analysis and design can be carried out exactly using z-transform techniques. The interested reader is referred to the bibliography in Appendix I.

The switched-capacitor circuits presented above are of the single-ended variety. In many applications, fully differential versions of these circuits are employed.

Summary

- A filter is a linear two-port network with a transfer function $T(s) = V_o(s)/V_i(s)$. For physical frequencies, the filter transmission is expressed as $T(j\omega) = |T(j\omega)|e^{j\phi(\omega)}$. The magnitude of transmission can be expressed in decibels using either the gain function $G(\omega) \equiv 20\log|T|$ or the attenuation function $A(\omega) \equiv -20\log|T|$.

- The transmission characteristics of a filter are specified in terms of the edges of the passband(s) and the stopband(s); the maximum allowed variation in passband transmission, A_{max} (dB); and the minimum attenuation required in the stopband, A_{min} (dB).

- The filter transfer function can be expressed as the ratio of two polynomials in s; the degree of the denominator polynomial, N, is the filter order. The N roots of the denominator polynomial are the poles.

- To obtain a highly selective response, the poles are complex and occur in conjugate pairs (except for one real pole when N is odd). The zeros are placed on the $j\omega$ axis in the stopband(s) including $\omega = 0$ and $\omega = \infty$.

- The Butterworth filter approximation provides a low-pass response that is maximally flat at $\omega = 0$. The transmission decreases monotonically as ω increases,

reaching 0 (infinite attenuation) at $\omega = \infty$, where all N transmission zeros lie. Eq. (14.19) gives $|T|$, where ϵ is given by Eq. (14.22) and the order N is determined using Eq. (14.23). The poles are found using the graphical construction of Fig. 14.15 and Eqs. (14.24) and (14.25), and the transfer function is given by Eq. (14.26).

- The Chebyshev filter approximation provides a low-pass response that is equiripple in the passband with the transmission decreasing monotonically in the stopband. All the transmission zeros are at $s = \infty$. Eq. (14.28) gives $|T|$ in the passband and Eq. (14.29) gives $|T|$ in the stopband, where ϵ is given by Eq. (14.31). The order N can be determined using Eq. (14.32). The poles are given by Eq. (14.33) and the transfer function by Eq. (14.34).
- Figures H.1 and H.2 (Appendix H, on the website) provide a summary of first-order filter functions and their realizations.
- Figure H.3 (Appendix H, on the website) provides the characteristics of seven special second-order filtering functions.
- The second-order LCR resonator of Fig. 14.18(a) realizes a pair of complex-conjugate poles with $\omega_0 = 1/\sqrt{LC}$ and $Q = \omega_0 CR$. This resonator can be used to realize the various special second-order filtering functions, as shown in Fig. 14.19.
- By replacing the inductor of an LCR resonator with a simulated inductance obtained using the Antoniou circuit of Fig. 14.20(a), the op amp–RC resonator of Fig. 14.21(b) is obtained. This resonator can be used to realize the various second-order filter functions as shown in Fig. 14.22 with the complete set of circuits shown in Fig. H.6 (Appendix H, on the website). The design equations for these circuits are given in Table H.1 (Appendix H, on the website).
- Biquads based on the two-integrator-loop topology are the most versatile and popular second-order filter realizations. There are two varieties: the KHN circuit of Fig. 14.24(a), which realizes the LP, BP, and HP functions simultaneously and can be combined with the output summing amplifier of Fig. 14.24(b) to realize the notch function; and the Tow–Thomas circuit of Fig. 14.25(b), which realizes the BP and LP functions simultaneously. Feedforward can be applied to the Tow–Thomas circuit to obtain the circuit of Fig. 14.26, which can be designed to realize any of the second-order functions (see Table H.2 in Appendix H on the website).
- Single-amplifier filters have the advantage of requiring low power but are limited to applications for which the pole Q is 5 or less. Examples of such circuits are shown in Fig. 14.27(a) (bandpass), Fig. 14.28(a) (high pass), and Fig. 14.29 (low pass). The full theory of single-amplifier biquads is provided on the website.
- Switched-capacitor (SC) filters are based on the principle that a capacitor C, periodically switched between two circuit nodes at a high rate, f_c, is equivalent to a resistance $R = 1/Cf_c$ connecting the two circuit nodes. SC filters can be fabricated in monolithic form using CMOS IC technology. The standard switched-capacitor integrators are shown in Fig. 14.31, and a biquad circuit using these integrators is shown in Fig. 14.32(b).

PROBLEMS

Section 14.1: Basic Concepts

14.1 The transfer function of a first-order low-pass filter (such as that realized by an RC circuit) can be expressed as $T(s) = \omega_0/(s+\omega_0)$, where ω_0 is the 3-dB frequency of the filter. Give in table form the values of $|T|$, ϕ, G, and A at $\omega = 0, 0.5\omega_0, \omega_0, 2\omega_0, 5\omega_0, 10\omega_0$, and $100\omega_0$.

14.2 A sinusoid with 1-V peak amplitude is applied at the input of a filter having the transfer function

$$T(s) = \frac{2\pi \times 10^4}{s + 2\pi \times 10^4}$$

Find the peak amplitude and the phase (relative to that of the input sinusoid) of the output sinusoid if the frequency of the

problems with blue numbers are considered essential; * = difficult problem; ** = more difficult; *** = very challenging
SIM = simulation; D = design problem; ● = see related video example

input sinusoid is (a) 1 kHz, (b) 10 kHz, (c) 100 kHz, and (d) 1 MHz.

***14.3** A filter has the transfer function $T(s) = 1/[(s+1)(s^2+s+1)]$. Show that $|T| = 1/\sqrt{1+\omega^6}$ and find an expression for its phase response $\phi(\omega)$. Calculate the values of $|T|$ for $\omega = 0.1, 0.2, 0.5, 1, 2, 5$, and 10 rad/s.

14.4 For the filter whose magnitude response is sketched (as the blue curve) in Fig. 14.3, find $|T|$ at $\omega = 0$, $\omega = \omega_p$, and $\omega = \omega_s$. $A_{max} = 0.5$ dB, and $A_{min} = 80$ dB.

D 14.5 A low-pass filter is required to pass all signals within its passband, extending from 0 to 5 kHz, with a transmission variation of at most 7% (i.e., the ratio of the maximum to minimum transmission in the passband should not exceed 1.07). The transmission in the stopband, which extends from 6 kHz to ∞, should not exceed 0.1% of the maximum passband transmission. What are the values of A_{max}, A_{min}, and the selectivity factor for this filter?

14.6 A low-pass filter is specified to have $f_p = 4$ kHz and a selectivity factor of 10. The specifications are just met by a first-order transfer function

$$T(s) = \frac{2\pi \times 10^4}{s + 2\pi \times 10^4}$$

What must A_{max} and A_{min} be?

14.7 A low-pass filter is specified to have $A_{max} = 1$ dB and $A_{min} = 10$ dB. We find that these specifications can be just met with a single-time-constant RC circuit having a time constant of 1 s and a dc transmission of unity. What must ω_p and ω_s of this filter be? What is the selectivity factor?

14.8 Sketch transmission specifications for a high-pass filter having a passband defined by $f \geq 5$ kHz and a stopband defined by $f \leq 4$ kHz. $A_{max} = 0.5$ dB, and $A_{min} = 70$ dB.

14.9 Sketch transmission specifications for a bandstop filter that is required to pass signals over the bands $0 \leq f \leq 12$ kHz and 20 kHz $\leq f \leq \infty$ with A_{max} of 0.4 dB. The stopband extends from $f = 15$ kHz to $f = 18$ kHz, with a minimum required attenuation of 60 dB.

Section 14.3: The Filter Transfer Function

14.10 A filter has transmission zeros at $\omega = 2$ rad/s and $\omega = \infty$. Its poles are at $s = -1$ and $s = -0.5 \pm j1.0$. The dc gain is unity. Find the filter order, type, and transfer function $T(s)$.

14.11 Consider a fifth-order filter whose poles are all at a radial distance from the origin of 10^5 rad/s. One pair of complex-conjugate poles is at 18° angles from the $j\omega$ axis, and the other pair is at 54° angles. Give ω_0 and Q of the poles and the transfer function in each of the following cases.

(a) The transmission zeros are all at $s = \infty$ and the dc gain is unity.

(b) The transmission zeros are all at $s = 0$ and the high-frequency gain is unity.

What type of filter results in each case?

14.12 Give the order and the transfer function of a bandpass filter that has the following poles and zeros:

Poles: a complex-conjugate pair with $\omega_0 = 1 \times 10^3$ rad/s and $Q = 1$, and another complex-conjugate pair with $\omega_0 = 1.2 \times 10^3$ rad/s and $Q = 12$.
Zeros: two at $\omega = 0$ and two at $\omega = \infty$.

The magnitude of the filter transmission at $\omega = 10^3$ rad/s is to be unity.

14.13 A second-order low-pass filter has poles at $-0.25 \pm j$ and a transmission zero at $\omega = 2.5$ rad/s. If the dc gain is unity, give the transfer function $T(s)$. What is the gain at ω approaching infinity?

14.14 Find the order N and the form of $T(s)$ of a bandpass filter having transmission zeros as follows: one at $\omega = 0$, one at $\omega = 10^3$ rad/s, one at 3×10^3 rad/s, and one at $\omega = \infty$. If this filter has a monotonically decreasing passband transmission with a peak at the center frequency of 2×10^3 rad/s, and equiripple response in the stopbands, sketch the shape of its $|T|$.

***14.15** Analyze the RLC network of Fig. P14.15 to determine its transfer function $V_o(s)/V_i(s)$ and hence its poles and zeros. (*Hint:* Begin the analysis at the output and work your way back to the input.)

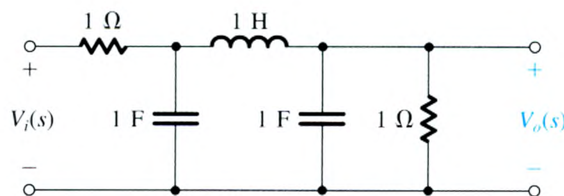


Figure P14.15

D 14.16 Design the circuit in Fig. 14.8 to realize a first-order op amp–RC low-pass filter having a 3-dB frequency of 10 kHz, a dc gain magnitude of 10, and an input resistance of 10 kΩ.

D 14.17 Use the circuit in Fig. E14.7 to design a first-order high-pass filter with a 3-dB frequency of 100 Hz, a high-frequency input resistance of 100 kΩ, and a high-frequency gain magnitude of unity.

14.18 Give the transfer function of a second-order maximally flat low-pass filter having $\omega_0 = 10^4$ rad/s and a dc gain of 10.

14.19 Give the transfer function of a second-order high-pass filter (two transmission zeros at $\omega = 0$) with $f_0 = 10$ kHz, $Q = 2.5$, and a high-frequency gain of unity.

14.20 Give the transfer function of a bandpass filter with a center frequency of 10^5 Hz, a 3-dB bandwidth of 2000 Hz, and a center-frequency gain of 5.

14.21 Give the transfer function of a low-pass notch filter with $\omega_0 = 10^4$ rad/s, $\omega_n = 1.25 \times 10^4$ rad/s, $Q = 2$, and a dc gain of unity. What is the asymptotic transmission at high frequencies?

D *14.22 By cascading a first-order op amp–RC low-pass circuit such as that in Fig. 14.8 with a first-order op amp–RC high-pass circuit such as that in Fig. E14.7, one can design a wideband bandpass filter. Provide such a design for the case in which the midband gain is 12 dB and the 3-dB bandwidth extends from 100 Hz to 100 kHz. Select appropriate component values under the constraints that no resistors higher than 100 kΩ are to be used and that the input resistance is to be as high as possible.

***14.23** (a) Derive an expression for the transfer function of the op amp–RC circuit shown in Fig. P14.23. Give expressions for the frequency of the transmission zero ω_Z, the frequency of the pole ω_P, the dc gain, and the high-frequency gain.

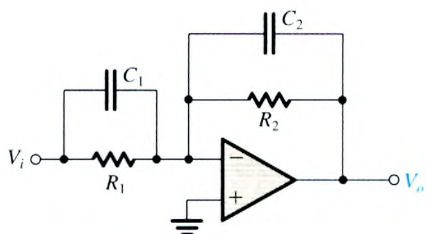


Figure P14.23

(b) Design the circuit to realize a first-order op amp–RC spectrum-shaping function with a transmission zero frequency of 100 Hz, a pole frequency of 10 kHz, and a dc gain magnitude of unity. The low-frequency input resistance is to be 10 kΩ. What high-frequency gain results? Sketch the magnitude of the transfer function versus frequency.

D *14.24 Derive $T(s)$ for the op amp–RC circuit in Fig. P14.24. Show that $|T(j\omega)| = 1$ and find $\phi(\omega)$. We wish to use this circuit as a variable phase shifter by adjusting R. If the input signal frequency is 5×10^3 rad/s and if $C = 10$ nF, find the values of R required to obtain phase shifts of –30°, –60°, –90°, –120°, and –150°. Note that this circuit is known as **all-pass** filter.

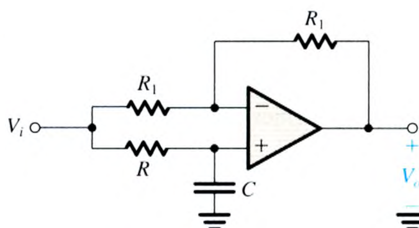


Figure P14.24

14.25 Find the transfer function of a second-order low-pass filter with $\omega_0 = 10^4$ rad/s, $Q = 2$, and dc gain = 1. At what frequency does $|T|$ peak? What is the peak transmission? (*Hint*: Use the information in Fig. 14.10.)

D *14.26 Use the information in Fig. 14.10 to obtain the transfer function of a second-order low-pass filter that just meets the specifications defined in Fig. 14.3 with $\omega_p = 1$ rad/s and $A_{max} = 3$ dB. Find ω_0 and Q. Also, if $\omega_s = 2$ rad/s, find the value of A_{min} obtained. Investigate two cases: (a) maximally flat passband ($Q = 1/\sqrt{2}$) and, (b), equiripple passband.

14.27 Find the transfer function of a second-order high-pass filter with a maximally flat passband response, a 3-dB frequency at $\omega = 1$ rad/s, and a high-frequency gain of unity. Give the location of the poles and zeros.

14.28 Find the transfer function of a second-order high-pass filter with poles at $-0.5 \pm j\sqrt{3}/2$ and a high-frequency gain of unity. What are ω_0 and Q of the poles?

D **14.29 (a) Show that $|T|$ of a second-order bandpass function is geometrically symmetrical around the center frequency ω_0. That is, the members of each pair of frequencies ω_1 and ω_2 for which $|T(j\omega_1)| = |T(j\omega_2)|$ are related by $\omega_1\omega_2 = \omega_0^2$.
(b) Find the transfer function of the second-order bandpass filter that meets specifications of the form in Fig. 14.4 where $\omega_{p1} = 8100$ rad/s, $\omega_{p2} = 10{,}000$ rad/s, and $A_{max} = 3$ dB. If $\omega_{s1} = 3000$ rad/s, find A_{min} and ω_{s2}.

D **14.30 (a) For a second-order notch function with $\omega_n = \omega_0$, show that for the attenuation to be greater than A dB over a frequency band BW_a, the value of Q must satisfy

$$Q \le \frac{\omega_0}{BW_a\sqrt{10^{A/10}-1}}$$

(*Hint*: First, show that any two frequencies, ω_1 and ω_2, at which $|T|$ is the same, are related by $\omega_1\omega_2 = \omega_0^2$).
(b) Use the result in (a) to find the transfer function of a notch filter that is required to eliminate a bothersome interference of 60-Hz frequency. Since the frequency of the interference is not stable, the filter should be designed to provide attenuation ≥ 20 dB over a 6-Hz band centered around 60 Hz. The dc transmission of the filter is to be unity.

14.31 Find the transfer function of a second-order high-pass notch filter for which $\omega_n = 1$ rad/s, $\omega_0 = 1.3$ rad/s, $Q = 3$, and the high-frequency asymptotic gain is unity. Sketch $|T(j\omega)|$ and give the value of the dc gain.

Section 14.3: Butterworth and Chebyshev Filters

14.32 Calculate the value of attenuation obtained at a frequency 2 times the 3-dB frequency of a seventh-order Butterworth filter.

14.33 Find the poles of a Butterworth filter having a 0.5-dB bandwidth of 10^3 rad/s and $N = 6$

D 14.34 Determine the order N of the Butterworth filter for which $A_{max} = 0.5$ dB, $A_{min} = 20$ dB, and the selectivity ratio $\omega_s/\omega_p = 1.7$. What is the actual value of minimum stopband attenuation realized? If A_{min} is to be exactly 20 dB, to what value can A_{max} be reduced?

14.35 Show that the order N of a Butterworth filter can be obtained from the approximate expression

$$N \ge \frac{A_{min} - 20\log\epsilon}{20\log(\omega_s/\omega_p)}$$

(*Hint:* Use Eq. (14.23) and neglect the unity term.)

D 14.36 Design a Butterworth filter that meets the following low-pass specifications: $f_p = 10$ kHz, $A_{max} = 3$ dB, $f_s = 20$ kHz, $A_{min} = 40$ dB, and dc gain = 1. Find N, the poles, and $T(s)$. What is the attenuation provided at 30 kHz?

14.37 Sketch the transfer function magnitude for a low-pass Chebyshev filter of (a) sixth order and (b) seventh order.

14.38 On the same diagram, sketch the magnitude of the transfer function of a Butterworth and a Chebyshev low-pass filter of fifth order and having the same ω_p and A_{max}. At the stopband edge, ω_s, which filter gives greater attenuation?

14.39 For a fifth-order Chebyshev filter with a unity transmission at dc and with a 1-dB passband ripple, find the attenuation realized at $f_s = 2f_p$.

14.40 Observe that Eq. (14.28) can be used to find the frequencies in the passband at which $|T|$ is at its peaks and at its valleys. (The peaks are reached when the $\cos^2[\ ]$ term is zero, and the valleys correspond to the $\cos^2[\ ]$ term equal to unity.) Find these frequencies for a fifth-order filter.

14.41 Contrast the attenuation provided by a sixth-order Chebyshev filter at $\omega_s = 2\omega_p$ to that provided by a Butterworth filter of equal order. For both, $A_{max} = 1$ dB. Sketch $|T|$ for both filters on the same axes.

D *14.42 Design a low-pass filter to meet the following specifications: $f_p = 3.4$ kHz, $A_{max} = 1$ dB, $f_s = 4$ kHz, $A_{min} = 35$ dB. The dc transmission is unity.

(a) Find the required order of Chebyshev filter. What is the excess (above 35 dB) stopband attenuation obtained?
(b) Find the poles and the transfer function.

Section 14.4: Second-Order Passive Filters Based on the LCR Resonator

14.43 Analyze the circuit in Fig. 14.18(c) to determine its transfer function $T(s) \equiv V_o(s)/V_i(s)$, and hence show that its poles are characterized by ω_0 and Q of Eqs. (14.39) and (14.40), respectively.

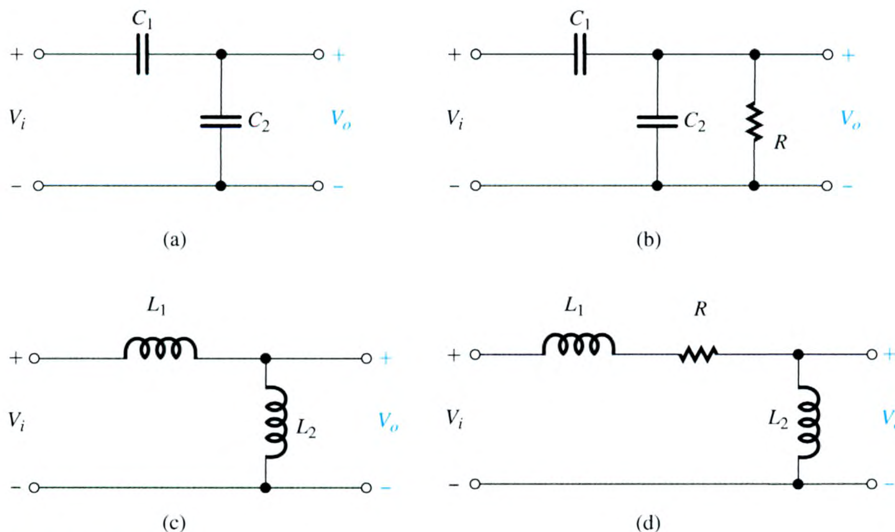


Figure P14.46

D 14.44 Design the LCR resonator of Fig. 14.18(a) to obtain poles with $\omega_0 = 2 \times 10^5$ rad/s and $Q = 4$. Use $R = 10$ kΩ.

14.45 For the LCR resonator of Fig. 14.18(a), find the change in ω_0 that results from

(a) increasing L by 1%
(b) increasing C by 1%
(c) decreasing R by 1%

14.46 For each of the circuits in Fig. P14.46, find the transmission as ω approaches zero and as ω approaches ∞, and hence find the transmission zeros.

14.47 For the resonator in Fig. 14.18(a), let node y be lifted from ground. If a signal source V_i is connected to node y and the output V_o is taken across the resonator, derive an expression for the transfer function V_o/V_i and show that it is a high pass with ω_0 and Q given by Eqs. (14.39) and (14.40), respectively. Design the high-pass circuit to obtain $\omega_0 = 10^6$ rad/s and $Q = 1$. Use $R = 1$ kΩ.

D *14.48 Show that the general notch circuit in Fig. 14.19(e) can be used to realize a low-pass notch (LPN) filter by making $L_2 = \infty$. (Recall that for the LPN case, $\omega_n > \omega_0$.) Design the circuit to obtain $\omega_0 = 10^5$ rad/s, $\omega_n = \sqrt{2} \times 10^5$ rad/s, and $Q = 2$. Use $R = 1$ kΩ. Find all the component values, the dc gain, and the high-frequency gain.

D 14.49 Show that the general notch circuit in Fig. 14.19(e) can be used to realize a high-pass notch (HPN) filter by making $C_2 = 0$. (Recall that for the HPN case, $\omega_n < \omega_0$.) Design the circuit for the case $\omega_0 = 10^5$ rad/s, $\omega_n = 10^5/\sqrt{2}$ rad/s, and $Q = 2$. Use $R = 1$ kΩ. Find all the component values, the dc gain, and the high-frequency gain.

D 14.50 Modify the bandpass circuit of Fig. 14.19(c) to change its center-frequency gain from 1 to 0.5 without changing ω_0 or Q.

14.51 Consider the LCR resonator of Fig. 14.18(a) with node x disconnected from ground and connected to an input signal source V_x, node y disconnected from ground and connected to another input signal source V_y, and node z disconnected from ground and connected to a third input signal source V_z. Use superposition to find the voltage that develops across the resonator, V_o, in terms of V_x, V_y, and V_z.

Section 14.5: Second-Order Active Filters Based on Inductance Simulation

D 14.52 Design the circuit of Fig. 14.20(a) (utilizing suitable component values) to realize an inductance of (a) 10 H, (b) 1.0 H, and (c) 0.1 H.

14.53 Figure P14.53 shows a generalized form of the Antoniou circuit of Fig. 14.20(a). Here, R_5 is eliminated

problems with blue numbers are considered essential; * = difficult problem; ** = more difficult; *** = very challenging
SIM = simulation; **D** = design problem; ● = see related video example

and the other four components are replaced by general impedances Z_1, Z_2, Z_3, and Z_4.

(a) With an impedance Z_5 connected between node 2 and ground, show that the input impedance looking into port 1 (i.e., between node 1 and ground) is

$$Z_{11} = \left(\frac{Z_1 Z_3}{Z_2 Z_4}\right) Z_5$$

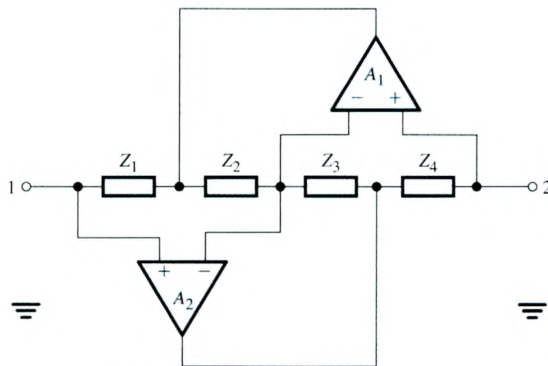


Figure P14.53

(b) From the symmetry of the circuit, show that if an impedance Z_6 is connected between terminal 1 and ground, the input impedance looking into port 2, which is between terminal 2 and ground, is given by

$$Z_{22} = \left(\frac{Z_2 Z_4}{Z_1 Z_3}\right) Z_6$$

(c) From the expressions above, observe that the two-port network in Fig. P14.53 acts as an "impedance transformer." Since by the appropriate choice of Z_1, Z_2, Z_3, and Z_4, the transformation ratio can be a general function of the complex frequency variable s, the circuit is known as a **generalized impedance converter**, or GIC.

***14.54** Consider the Antoniou circuit of Fig. 14.20(a) with R_5 eliminated, a capacitor C_6 connected between node 1 and ground, and a voltage source V_2 connected to node 2. Show that the input impedance seen by V_2 is $R_2/s^2C_4C_6R_1R_3$. How does this impedance behave for physical frequencies ($s = j\omega$)? (This impedance is known as a **frequency-dependent negative resistance**, or FDNR.)

***14.55** Starting from first principles and assuming ideal op amps, derive the transfer function of the circuit in Fig. 14.22(a).

D *14.56 Design a fifth-order Butterworth filter having a 3-dB bandwidth of 10^4 rad/s and a dc gain of 10 V/V. Use a cascade of two circuits of the type shown in Fig. 14.22(a) and a first-order op amp–RC circuit. Select appropriate component values.

D 14.57 Give the circuit for a high-pass filter based on the resonator of Fig. 14.21(b). Design the circuit to obtain $\omega_0 = 10^3$ rad/s, $Q = 5$, and a high-frequency gain of unity.

D 14.58 Show that the circuit in Fig. 14.22(c) can be used to realize a high-pass notch filter ($\omega_n < \omega_0$) by selecting $C_{62} = 0$. Design the circuit to obtain $\omega_0 = 10^5$ rad/s, $\omega_n = 0.8 \times 10^5$ rad/s, $Q = 5$, and a high-frequency gain of 5.

D 14.59 Use the circuit in Fig. 14.22(b) to design a bandpass filter with $f_0 = 100$ kHz and a 3-dB bandwidth of 1 kHz. The gain at f_0 is required be 10 V/V.

D **14.60 Design a third-order low-pass filter whose $|T|$ is equiripple in both the passband and the stopband (in the manner shown in Fig. 14.3, except that the response shown is for $N = 5$). The filter passband extends from $\omega = 0$ to $\omega = 1$ rad/s, and the passband transmission varies between 1 and 0.9. The stopband edge is at $\omega = 1.2$ rad/s. The following transfer function was obtained using filter-design tables:

$$T(s) = \frac{0.4508\left(s^2 + 1.6996\right)}{(s + 0.7294)(s^2 + s0.2786 + 1.0504)}$$

The actual filter realized is to have $\omega_p = 10^5$ rad/s.

(a) Obtain the transfer function of the actual filter by replacing s by $s/10^5$.

(b) Realize this filter as the cascade connection of a first-order LP op amp–RC circuit of the type shown in Fig. 14.8 and a second-order LPN circuit of the type shown in Fig. 14.22(c) with $R_{52} = \infty$. Each section is to have a dc gain of unity. Select appropriate component values. (*Note:* A filter with an equiripple response in both the passband and the stopband is known as an **elliptic filter**.)

Section 14.6: Second-Order Active Filters Based on the Two-Integrator Loop

D 14.61 Design the KHN circuit of Fig. 14.24(a) to realize a bandpass filter with a center frequency of 3 kHz and a 3-dB bandwidth of 50 Hz. Use 10-nF capacitors. Give the complete circuit and specify all component values. What value of center-frequency gain is obtained?

14.62 (a) Consider the KHN biquad in Fig. 14.24(a), with the three outputs V_{hp}, V_{bp}, and V_{lp} summed by the summer in Fig. 14.24(b). Using the transfer functions in Eqs. (14.59),

(14.62), and (14.63), show that the transfer function realized at the output V_o is

$$\frac{V_o}{V_i} = -K\,\frac{(R_F/R_H)s^2 - s(R_F/R_B)\omega_0 + (R_F/R_L)\omega_0^2}{s^2 + s(\omega_0/Q) + \omega_0^2}$$

(b) If a notch filter with a notch frequency ω_n and a dc gain G is required, find the required values of R_H, R_B, R_L, and R_F.

14.63 Use the Tow–Thomas circuit in Fig. 14.25(b) to design a low-pass filter with $f_0 = 10$ kHz, $Q = 20$, and dc gain = 10 V/V, and input resistance = 1 kΩ. Specify all component values.

D 14.64 Use the Tow–Thomas biquad of Fig. 14.26 to realize a high-pass notch filter with a notch frequency ω_n and high-frequency gain of G. If a convenient value is selected for C, give expressions for the required values of R, R_1, C_1, R_2, and R_3 in terms of ω_0, ω_n, C, and G.

14.65 Starting from first principles, derive the transfer function V_o/V_i of the circuit in Fig. 14.26 and verify that it is the expression given by Eq. (14.70).

D 14.66 Design the circuit of Fig. 14.26 to realize a low-pass notch filter with $\omega_0 = 10^5$ rad/s, $Q = 10$, dc gain = 1, and $\omega_n = 1.3 \times 10^5$ rad/s. Use $C = 10$ nF and $r = 20$ kΩ.

D **14.67 Repeat Problem 14.60 using the Tow–Thomas biquad of Fig. 14.26 to realize the second-order section in the cascade.

Section 14.7: Second-Order Active Filters Using a Single Op Amp

D 14.68 Use the circuit in Fig. 14.27(a) to realize a bandpass filter with a center frequency of 10 kHz and a 3-dB bandwidth of 2.5 kHz. Give the values of all components and specify the center-frequency gain obtained.

D *14.69 Consider the bandpass circuit shown in Fig. 14.27(a). Let $C_1 = C_2 = C$, $R_3 = R$, $R_4 = R/4Q^2$, and $CR = 2Q/\omega_0$. Disconnect the positive input terminal of the op amp from ground and apply V_i through a voltage divider R_1, R_2 to the positive input terminal as well as through R_4 as before. Analyze the circuit to find its transfer function V_o/V_i. Find the ratio R_2/R_1 so that the circuit realizes a notch function.

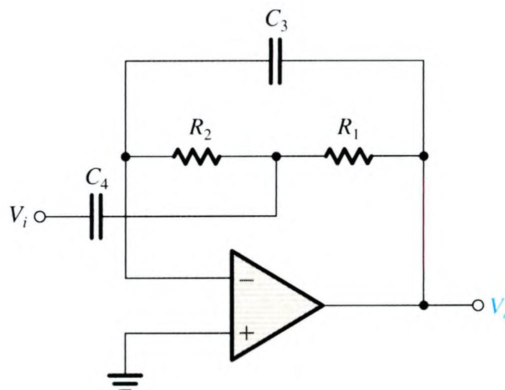


Figure P14.70

14.70 Analyze the circuit in Fig. P14.70 to determine its transfer function $T(s) \equiv V_o/V_i$. Show that $T(s)$ is that of a second-order bandpass filter and find ω_0 and Q. For $R_1 = R_2 = R$, $C_4 = C$, and $C_3 = C/m$, find CR and m in terms of ω_0 and Q. What center-frequency gain is obtained?

14.71 Use a cascade of the low-pass circuit in Fig. 14.29 and the high-pass circuit in Fig. 14.28(a) to realize a wide-band bandpass filter with 3-dB frequencies of 1 kHz and 100 kHz. The response of both circuits should be maximally flat; that is, each should have a Q of 0.707. Use capacitances in the nanofarad range. Sketch and label a Bode plot for the gain magnitude.

D *14.72 Design a fifth-order Butterworth low-pass filter that has a 3-dB bandwidth of 10 kHz and a dc gain of unity. Use the cascade connection of two circuits of the type shown in Fig. 14.29 and a first-order low-pass circuit (Fig. 14.8). Use a 10-kΩ value for all resistors.

Section 14.8: Switched-Capacitor Filters

14.73 For the switched-capacitor input circuit of Fig. 14.30(b), in which a clock frequency of 100 kHz is used, what input resistances correspond to capacitance C_1 values of 0.1 pF, 0.5 pF, 1 pF, 5 pF, and 10 pF?

14.74 For a dc voltage of 1 V applied to the input of the circuit of Fig. 14.30(b), in which C_1 is 1 pF, what charge is transferred for each cycle of the two-phase clock? For a 200-kHz clock, what is the average current drawn from the input source? For a feedback capacitance of 10 pF, what change would you expect in the output for each cycle of the clock? For an amplifier that saturates at ± 1 V and the feedback capacitor initially discharged, how many clock cycles would it take to saturate the amplifier? What is the average slope of the staircase output voltage produced?

D 14.75 Repeat Exercise 14.24 for a clock frequency of 500 kHz.

D 14.76 Repeat Exercise 14.24 for $Q = 40$.

D 14.77 Design the circuit of Fig. 14.32(b) to realize, at the output of the second (noninverting) integrator, a maximally flat low-pass function with $\omega_{3dB} = 10^4$ rad/s and unity dc gain. Use a clock frequency $f_c = 200$ kHz and select $C_1 = C_2 = 2$ pF. Give the values of C_3, C_4, C_5, and C_6. (*Hint:* For a maximally flat response, $Q = 1/\sqrt{2}$ and $\omega_{3dB} = \omega_0$.)

CHAPTER 15

Oscillators

IN THIS CHAPTER YOU WILL LEARN

- That an oscillator circuit that generates sine waves can be implemented by connecting a frequency-selective network (a filter) in the positive-feedback path of an amplifier.
- The conditions under which sustained oscillations are obtained and the frequency of those oscillations.
- How to control the amplitude of the sine wave obtained in an oscillator using a nonlinear circuit.
- A variety of sine-wave oscillator circuits.
- How to use an op amp to design a bistable circuit that can be connected in a feedback loop with an op-amp integrator to implement a generator of square and triangular waveforms.

Introduction

In the design of electronic systems, we often need signals with prescribed standard waveforms, such as sinusoidal, square, triangular, or pulse. Systems that require standard signals include computer and control systems, where clock pulses are needed for, among other things, timing; communication systems, where signals of a variety of waveforms are used as information carriers; and test and measurement systems, where signals, again of a variety of waveforms, are used to test and characterize electronic devices and circuits. In this chapter we study signal-generator circuits.

The signal-generator or oscillator circuits we will study are collectively capable of providing signals with frequencies in the range of hertz to hundreds of gigahertz. While some can be fabricated on chips, others use discrete components. Examples of commonly encountered oscillators include the microprocessor clock generator (fabricated on chip utilizing the ring oscillator studied in Section 18.4 with frequencies in the several-gigahertz range); the carrier-waveform generator in wireless transceivers (on chip, up to the hundreds-of-gigahertz range); the oscillator in an electronic watch (using a quartz crystal with a frequency of 2^{15} Hz); and the variable-frequency function generator in the electronics lab (utilizing a discrete circuit with frequency in the hertz to megahertz range).

There are two distinctly different approaches to designing oscillator circuits. The first, used to generate sine waves and studied in Sections 15.1 to 15.3, uses a **positive-feedback loop** consisting of an amplifier and an RC or LC **frequency-selective network**. While the frequency of the generated sine wave is determined by the frequency-selective network, the amplitude is set using a nonlinear mechanism, implemented either with a separate circuit or using the nonlinearities of the amplifying device itself. In spite of this, these circuits, which generate sine waves utilizing resonance phenomena, are known as **linear oscillators**. The reason behind this name will be explained later. The second approach uses a circuit building block known as a **bistable multivibrator** together with an RC circuit in a feedback loop to generate square waves. If the RC circuit is replaced with an op-amp integrator, triangular waves can be obtained as well. For our purposes here, the bistable multivibrator is realized using an op amp connected in a positive-feedback loop. This type of oscillator circuit is known as a **nonlinear oscillator** or, more commonly, as a **function generator**. We will study bistable circuits and function generators in Section 15.4.

15.1 Basic Principles of Sinusoidal Oscillators

In this section, we study the basic principles of the design of linear sine-wave oscillators. In spite of the name *linear oscillator,* we have to use some form of nonlinearity to control the amplitude of the output sine wave. In fact, all oscillators are essentially nonlinear circuits. This complicates the analysis and design of oscillators: no longer can we apply transform (s-plane) methods directly. Nevertheless, engineers have developed techniques to design sinusoidal oscillators in two steps: The first step is a linear one, using frequency-domain methods of feedback circuit analysis; the second step provides a nonlinear mechanism for amplitude control.

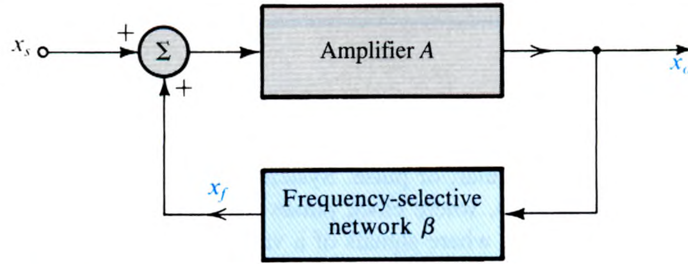


Figure 15.1 The basic structure of a sinusoidal oscillator. A positive-feedback loop is formed by an amplifier and a frequency-selective network. In an actual oscillator circuit, no input signal will be present; here we are showing an input signal x_s to help explain the principle of operation.

15.1.1 The Oscillator Feedback Loop

The basic structure of a sinusoidal oscillator consists of an amplifier and a frequency-selective network connected in a **positive-feedback loop**, like the one shown in block diagram form in Fig. 15.1. Although no input signal will be present in an actual oscillator circuit, we have included an input signal here to help explain the principle of operation. It is important to note that unlike the negative-feedback loop of Fig. 11.1, here the feedback signal x_f is summed with a *positive* sign. Thus the gain-with-feedback is given by

$$A_f(s) = \frac{A(s)}{1 - A(s)\beta(s)} \tag{15.1}$$

where we note the negative sign in the denominator. The loop gain $L(s)$ is given by

$$L(s) \equiv A(s)\beta(s) \tag{15.2}$$

and the characteristic equation is

$$1 - L(s) = 0 \tag{15.3}$$

15.1.2 The Oscillation Criterion

If at a specific frequency ω_0 the loop gain $A\beta$ is equal to unity, it follows from Eq. (15.1) that A_f will be infinite. In other words, at this frequency the circuit will have a finite output for zero input signal. Such a circuit is by definition an oscillator. Thus the condition for the feedback loop of Fig. 15.1 to provide sinusoidal oscillations of frequency ω_0 is

$$L(j\omega_0) \equiv A(j\omega_0)\beta(j\omega_0) = 1 \tag{15.4}$$

That is, *at ω_0 the phase of the loop gain should be zero, and the magnitude of the loop gain should be unity.* This is known as the **Barkhausen criterion**. Note that for the circuit to oscillate at one frequency, the oscillation criterion should be satisfied only at one frequency (i.e., ω_0); otherwise the resulting waveform will not be a simple sinusoid.

We can get a feel for the Barkhausen criterion by considering once more the feedback loop of Fig. 15.1. For this loop to *produce* and *sustain* an output x_o with no input applied ($x_s = 0$), the feedback signal x_f,

$$x_f = \beta x_o$$

should be sufficiently large that when multiplied by A it produces x_o, that is,

$$Ax_f = x_o$$

or

$$A\beta x_o = x_o$$

which results in

$$A\beta = 1$$

Note that the *frequency of oscillation* ω_0 is determined solely by the phase characteristics of the feedback loop; the loop oscillates at the frequency for which the phase is zero (or, equivalently, 360°).

An alternative approach for studying oscillator circuits is to examine the circuit poles, which are the roots of the **characteristic equation** (Eq. 15.3). For the circuit to produce **sustained oscillations** at a frequency ω_0 the characteristic equation has to have roots at $s = \pm j\omega_0$. Thus $1 - A(s)\beta(s)$ must have a factor of the form $s^2 + \omega_0^2$.

EXERCISE

15.1 Consider a sinusoidal oscillator formed by connecting an amplifier with a gain of 2 and a second-order bandpass filter in a feedback loop. Find the pole frequency and the center-frequency gain of the filter needed to produce sustained oscillations at 1 kHz.
Ans. 1 kHz; 0.5

15.1.3 Analysis of Oscillator Circuits

For a given oscillator circuit, it takes three steps to determine the frequency of oscillation and the condition for the oscillations to start:

1. Break the feedback loop to determine the loop gain $A(s)\beta(s)$. This step is similar to the one we used in Section 11.2 when we analyzed negative-feedback amplifiers.
2. Find the oscillation frequency ω_0 as the frequency for which the phase angle of $A(j\omega)\beta(j\omega)$ is zero or, equivalently, 360°.
3. Find the condition for the oscillations to start from

$$|A(j\omega_0)\beta(j\omega_0)| \geq 1$$

Note that making the magnitude of the loop gain slightly greater than one ensures that oscillations will start.

Example 15.1

Figure 15.2(a) shows a sinusoidal oscillator formed by placing a second-order LCR bandpass filter [see Fig. 14.19(c)] in the feedback path of a positive-gain amplifier. Find the frequency of oscillation ω_0, and the condition for oscillations to start. Assume an ideal op amp.

Solution

To obtain the loop gain, we break the positive-feedback loop at the positive input terminal of the op amp where the input impedance is infinite, apply an input voltage V_i, and find the returned voltage V_r. This results in the circuit in Fig. 15.2(b). Since the op amp is ideal, we can find its output voltage V_o and hence $A(s)$ as

$$A(s) \equiv \frac{V_o}{V_i} = 1 + \frac{r_2}{r_1}$$

We can find the transfer function of the frequency-selective network $\beta(s) \equiv V_r/V_o$ either by direct analysis or using Eq. (14.43) as

$$\beta(s) \equiv \frac{V_r}{V_o} = \frac{s\dfrac{1}{CR}}{s^2 + s\dfrac{1}{CR} + \dfrac{1}{LC}}$$

While we have found it easy to determine $A(s)$ and $\beta(s)$ separately, our interest is in fact in their product, the loop gain $A(s)\,\beta(s)$,

$$A(s)\beta(s) = \frac{s\dfrac{1}{CR}\left(1 + \dfrac{r_2}{r_1}\right)}{s^2 + s\dfrac{1}{CR} + \dfrac{1}{LC}}$$

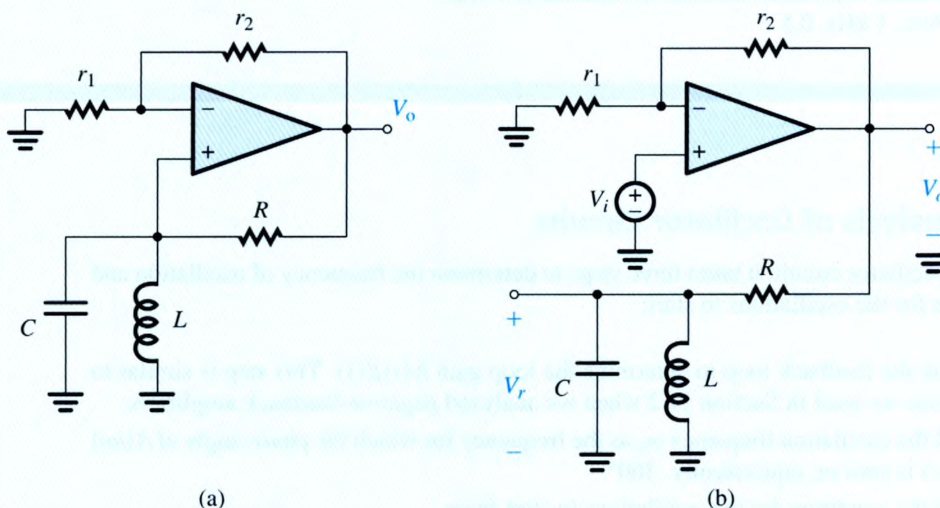


Figure 15.2 **(a)** An oscillator formed by connecting a positive-gain amplifier in a feedback loop with a bandpass RLC circuit. **(b)** Breaking the feedback loop at the input of the op amp to determine $A(s) \equiv V_o(s)/V_i(s)$ and $\beta(s) \equiv V_r(s)/V_o(s)$, and hence the loop gain $A(s)\beta(s)$.

Substituting $s = j\omega$,

$$A(j\omega)\beta(j\omega) = \frac{j\dfrac{\omega}{CR}\left(1+\dfrac{r_2}{r_1}\right)}{\left(-\omega^2+\dfrac{1}{LC}\right)+j\dfrac{\omega}{CR}}$$

From this expression we see that the phase angle of $A(j\omega)\ \beta(j\omega)$ will be zero at the value of ω that makes the real part of the denominator zero, thus

$$\omega_0 = 1/\sqrt{LC}$$

At this frequency, the magnitude of the loop gain is given by

$$|A(j\omega_0)\beta(j\omega_0)| = 1+\frac{r_2}{r_1}$$

Thus oscillation will start for

$$r_2/r_1 \geq 0$$

While $r_2 = 0$ is theoretically sufficient for sustained oscillations, we usually design for $r_2/r_1 > 0$ to ensure that oscillations start. This point will be explained in greater detail in the next section.

Before leaving this example, we observe that the results could have been obtained by inspection of the circuit in Fig. 15.2(a). Since the phase angle of the amplifier gain is zero, we examine the LCR bandpass circuit to determine the frequency at which its phase is zero. This in turn is the frequency at which the LC tank has an infinite impedance, which is the resonance frequency $\omega_0 = 1/\sqrt{LC}$. At this frequency the transmission of the bandpass LCR circuit is unity. Thus for oscillation to start, the amplifier gain must be greater than or equal to unity.

An Alternative Analysis Approach There is a simple alternative for analyzing oscillator circuits without having to break the feedback loop: We assume that the circuit is oscillating and thus has voltage and current signals at the oscillation frequency ω_0. Then we analyze the circuit in the usual manner and reduce the equations to a single equation in terms of a single voltage or current variable. Since the voltage or current quantity is not zero (because we have assumed the circuit to be oscillating), we can divide by the variable and thus eliminate it. The equation can then be manipulated to the form

$$D(s) = 0$$

where $D(s)$ is a polynomial in s. Substituting $s = j\omega$, we then equate the real and imaginary parts of $D(j\omega)$ to zero. One of the resulting equations yields ω_0 and the other provides the condition for sustained oscillation. We will use this method various times in the next sections. The method will be illustrated by the following example.

Example 15.2

Use the alternative analysis method described above to find the frequency of oscillation ω_0 and the condition for sustained oscillation of the circuit in Fig. 15.2(a).

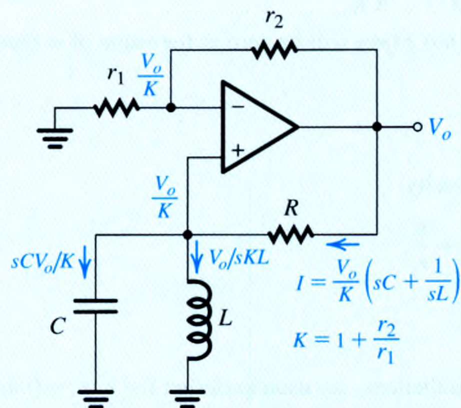


Figure 15.3 Analysis of the oscillator circuit of Fig. 15.2(a).

Solution

Figure 15.3 shows the circuit with some of the analysis already indicated. Writing a loop equation gives us

$$V_o = IR + \frac{V_o}{K} = \frac{V_o}{K}R\left(sC + \frac{1}{sL}\right) + \frac{V_o}{K} \tag{15.5}$$

where

$$K = 1 + \frac{r_2}{r_1}$$

Multiplying Eq. (15.5) by (sK/CR) and dividing by V_o and rearranging gives

$$s^2 + s\frac{1}{CR}(1-K) + \frac{1}{LC} = 0$$

Substituting $K = 1 + \frac{r_2}{r_1}$ gives

$$s^2 - s\frac{1}{CR}\frac{r_2}{r_1} + \frac{1}{LC} = 0 \tag{15.6}$$

For $s = j\omega$, we have

$$\left(-\omega^2 + \frac{1}{LC}\right) - j\omega\frac{1}{CR}\frac{r_2}{r_1} = 0$$

$$R_e[D(j\omega)] = -\omega^2 + \frac{1}{LC}$$

$$\text{Im}[D(j\omega)] = -\frac{\omega}{CR}\frac{r_2}{r_1}$$

For sustained oscillation, both the real part and the imaginary part must be zero at the frequency of oscillation, thus

$$R_e = 0 \Rightarrow \omega_0^2 = \frac{1}{LC} \Rightarrow \omega_0 = \frac{1}{\sqrt{LC}}$$

$$\text{Im} = 0 \Rightarrow \frac{r_2}{r_1} = 0 \Rightarrow r_2 = 0$$

To ensure that oscillations start, we usually design for

$$r_2 > 0$$

that is, a small resistance r_2 is used.

EXERCISES

15.2 For the oscillator circuit in Fig. 15.2(a), let $r_1 = 10\ \text{k}\Omega$, $r_2 = 100\ \Omega$, $R = 10\ \text{k}\Omega$, $C = 10$ nF, and $L = 0.1$ mH. Find the frequency of oscillation ω_0. Is the condition of oscillation satisfied?
Ans. 10^6 rad/s; yes, since $r_2 > 0$.

15.3 For the oscillator circuit of Fig. 15.2(a), find the percentage change in ω_0 that results from (a) L increasing by 1%, (b) C increasing by 1%, and (c) R increasing by 1%.
Ans. (a) −0.5%; (b) −0.5%; (c) 0%.

15.1.4 Nonlinear Amplitude Control

The oscillation condition, the Barkhausen criterion discussed in Section 15.1.2, guarantees sustained oscillations in a mathematical sense. It is well known, however, that the parameters of any physical system cannot be held constant for very long. Suppose we work hard to make $|A\beta| = 1$ at $\omega = \omega_0$, and then the temperature changes and $|A\beta|$ becomes slightly less than unity: obviously, oscillations will cease. On the other hand, if $|A\beta|$ exceeds unity, oscillations will grow in amplitude. Therefore we need a way to keep $|A\beta|$ equal to unity *at the desired value of output amplitude*. We do this by providing a nonlinear circuit for gain control.

Basically, the function of the gain-control mechanism is as follows: First, to ensure that oscillations will start, we design the circuit so that $|A\beta|$ is greater than unity. This corresponds to designing the circuit so that the poles are in the right half of the s plane. Thus as the power supply is turned on, oscillations will grow in amplitude. The increase in amplitude will be detected by the nonlinear circuit, which in turn will cause the gain to decrease. Eventually, when the amplitude reaches the desired level, the gain-control circuit makes $|A\beta|$ exactly

unity. In other words, the poles will be "pulled back" to the $j\omega$ axis. This action will cause the circuit to sustain oscillations at this desired amplitude. If, for some reason, the loop gain is reduced below unity, the amplitude of the sine wave will diminish. This will be detected by the nonlinear network, which will cause the loop gain to increase to exactly unity.

A straightforward implementation of the gain-control mechanism is obtained by using a nonlinear element, such as a diode, in place of one of the resistors that determine the amplifier gain. Since the resistance of the diode decreases as the amplitude of the signal across it increases, by appropriately placing the diode we can arrange that the gain of the amplifier decreases as the amplitude of the output sinusoid increases. An example is shown in Fig. 15.4, where two diodes are placed with opposite polarities across the resistance R_2 in the feedback path of the op amp. When the amplitude of the output signal is small, D_1 and D_2 will be cut off, and the feedback resistance will be $(R_2 + R_3)$. As the amplitude increases, D_1 and D_2 will conduct, and their incremental resistances will shunt R_2, causing the effective feedback resistance to decrease and the amplifier gain to correspondingly decrease.

An alternative approach to implementing amplitude control in oscillators uses a limiter (see Section 4.1.3). An example is shown in Fig. 15.5. Here, we have inserted a simple diode limiter, formed by diodes D_1 and D_2 together with resistor r_3, in the feedback loop of the oscillator of Fig. 15.2(a). As the amplitude of the oscillations grows, the limiter operates and provides an output v_L that approximates a square wave of frequency ω_0 and an amplitude

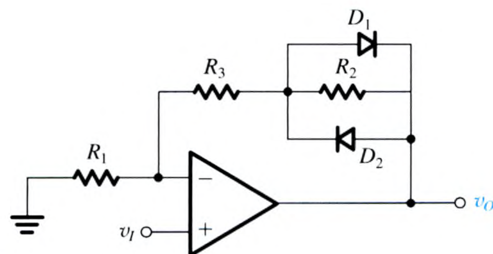


Figure 15.4 By placing two diodes in opposite directions across part of the feedback resistance, the amplifier gain v_O/v_I is made to decrease as the amplitude of v_O increases.

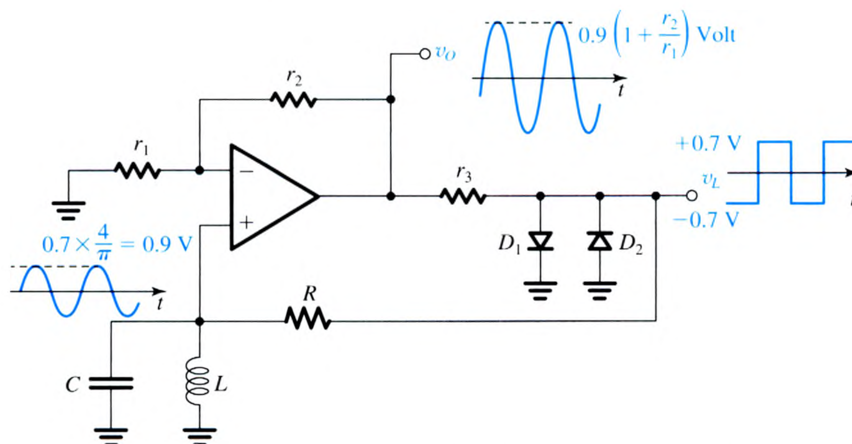


Figure 15.5 The oscillator circuit of Fig. 15.2 with a limiter inserted in the feedback loop for amplitude control.

whose precise value depends on the i-v characteristic of the diode and the value of r_3. In Fig. 15.5 we assume the output levels of the limiter to be ± 0.7 V. This square-wave signal is applied to the RLC bandpass filter. Assuming the filter is sufficiently selective (i.e., having a high Q-factor), its output across the LC tank will consist mostly of the fundamental-frequency component of the square wave; that is, a sinusoid with frequency ω_0. From the Fourier series of a square wave of amplitude V, we know that the fundamental frequency has an amplitude of $(4/\pi)V$. Thus, for our case, the amplitude of the sinusoid across the LC circuit will be $(4 \times 0.7/\pi)$ or approximately 0.9 V, and that at the amplifier output will be $\left[0.9(1 + r_2/r_1)\right]$ volts.

The circuit in Fig. 15.5 is easy to design and adjust since the frequency ω_0 is determined by the center-frequency of the filter, the purity of the sine wave is determined by the Q factor of the filter ($Q = \omega_0 CR$), and the amplitude of the output sine wave is determined by amplifier gain $(1 + r_2/r_1)$.

EXERCISE

15.4 For the oscillator circuit in Fig 15.5, find the value of (r_2/r_1) to obtain an output amplitude of 5 V.
Ans. 4.6

15.2 Op Amp–RC Oscillator Circuits

In this section we study some practical oscillator circuits using op amps and RC networks. These circuits are usually assembled on printed-circuit boards; they operate at frequencies ranging from very low to the low megahertz.

15.2.1 The Wien-Bridge Oscillator

One of the simplest oscillator circuits is based on the Wien bridge (see the highlighted note on page 1077). Figure 15.6 shows a Wien-bridge oscillator without the nonlinear gain-control network. The circuit consists of an op amp connected in the noninverting configuration, with a closed-loop gain of $1 + R_2/R_1$. In the feedback path of this positive-gain amplifier, an RC network having a bandpass tranfer function is connected. In this regard, the circuit is very similar to that in Fig. 15.2(a) except that here the bandpass filter is implemented with an RC circuit. The loop gain can be easily obtained by multiplying the transfer function $V_a(s)/V_o(s)$ of the feedback network by the amplifier gain,

$$L(s) = \left[1 + \frac{R_2}{R_1}\right]\frac{Z_p}{Z_p + Z_s}$$

$$= \frac{1 + R_2/R_1}{1 + Z_s Y_p}$$

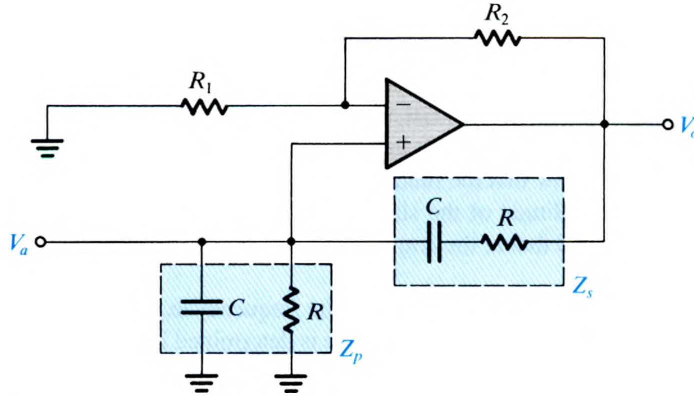


Figure 15.6 A Wien-bridge oscillator without amplitude stabilization.

Thus,

$$L(s) = \frac{1 + R_2/R_1}{3 + sCR + 1/sCR} \tag{15.7}$$

Substituting $s = j\omega$ results in

$$L(j\omega) = \frac{1 + R_2/R_1}{3 + j(\omega CR - 1/\omega CR)} \tag{15.8}$$

The loop gain will be a real number (i.e., the phase will be zero) at one frequency given by

$$\omega_0 CR = \frac{1}{\omega_0 CR}$$

That is,

$$\omega_0 = 1/CR \tag{15.9}$$

Oscillations will start at this frequency if the loop gain is at least one, which we can achieve by selecting

$$R_2/R_1 = 2 \tag{15.10}$$

To ensure that oscillations will start, we choose R_2/R_1 slightly greater than 2. We can easily verify that if $R_2/R_1 = 2 + \delta$, where δ is a small number, the roots of the characteristic equation $1 - L(s) = 0$ will be in the right half of the s plane.

The amplitude of oscillation can be set and stabilized by using a nonlinear control network. Two different implementations of the amplitude-controlling function are shown in Figs. 15.7 and 15.8. The circuit in Fig. 15.7 uses a symmetrical feedback limiter formed by diodes D_1 and D_2 together with resistors R_3, R_4, R_5, and R_6. The limiter operates in the following manner: The positive peak of the output voltage v_O will be limited to the value that causes the voltage at node b to exceed the voltage v_1 (which is about $1/3v_O$), by an amount sufficient to cause diode D_2 conducts. This will clamp the positive peak to a value determined by R_5, R_6, and the negative power supply V_{SS}. To be specific, the value of

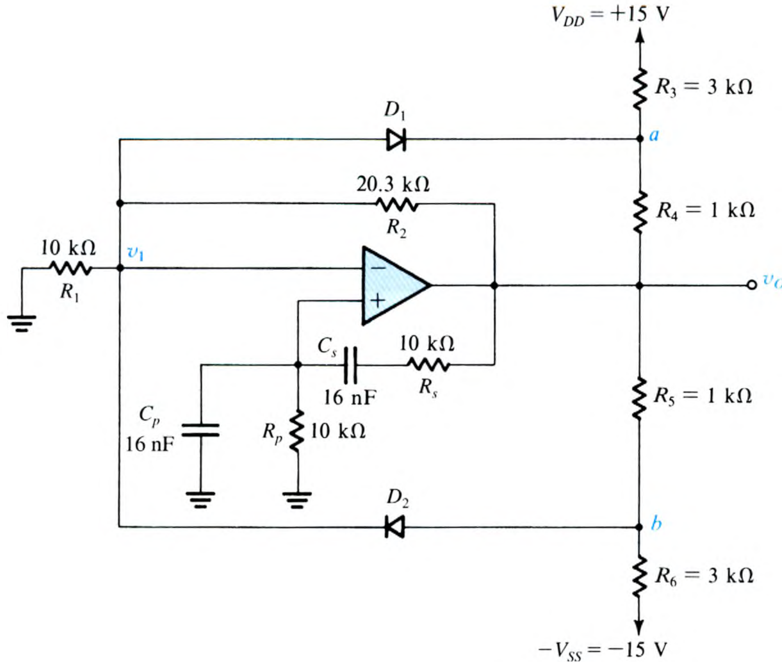


Figure 15.7 A Wien-bridge oscillator with a limiter used for amplitude control.

the positive output peak $\hat{v}_{O+}$ can be calculated by setting $v_b = v_1 + V_{D2} = 1/3\hat{v}_{O+} + V_D$ and writing a node equation at node b while neglecting the current through D_2. The result is:

$$\hat{v}_{O+} = \left[\left(\frac{R_5}{R_6}\right)V_{SS} + \left(1 + \frac{R_5}{R_6}\right)V_D\right] \Big/ \left(\frac{2}{3} - \frac{1}{3}\frac{R_5}{R_6}\right) \tag{15.11}$$

Similarly, the negative peak of the output sine wave will be clamped to the value that causes diode D_1 to conduct. The value of the negative peak $\hat{v}_{O-}$ can be determined by setting $v_a = v_1 - V_{D1} = 1/3\hat{v}_{O-} - V_D$ and writing an equation at node a while neglecting the current through D_1. The result is:

$$\hat{v}_{O-} = -\left[\left(\frac{R_4}{R_3}\right)V_{DD} + \left(1 + \frac{R_4}{R_3}\right)V_D\right] \Big/ \left(\frac{2}{3} - \frac{1}{3}\frac{R_4}{R_3}\right) \tag{15.12}$$

THE WIEN-BRIDGE OSCILLATOR

The Wien bridge consisting of four resistors and two capacitors was invented in 1891 by Max Wien, a Prussian physicist, for inductance measurement. Much later, William Hewlett (cofounder in 1939 of Hewlett-Packard), while working toward his master's degree at Stanford University, realized the importance of placing part of the Wien bridge in a positive-feedback loop to form what was called the Wien-bridge oscillator. The first product in 1939 of the new Hewlett-Packard Company was the HP200A, a flexible, precision sine-wave generator using vacuum tubes to implement the amplifier and a tungsten lamp to control the loop gain and thus the amplitude of the sine wave.

Note that in deriving Eqs. (15.11) and (15.12) we assumed $V_{D1} = V_{D2} = V_D$ and $v_1 = 1/3\ v_O$. To obtain a symmetrical output waveform, Eqs. (15.11) and (15.12) indicate that we choose $R_3 = R_6$, $R_4 = R_5$, and $V_{SS} = V_{DD}$. A final point: Note that the sine wave v_1 will have less distortion than v_O. This is because v_1 is equal to the signal at the positive input terminal of the op amp, which is a filtered version of v_O. On the negative side, the node at which v_1 appears is a high-impedance node and thus is not a convenient output node.

EXERCISE

15.5 For the circuit in Fig. 15.7: (a) Disregarding the limiter circuit, find the location of the closed-loop poles. (b) Find the frequency of oscillation. (c) With the limiter in place, find the amplitude of the output sine wave (assume that the diode drop is 0.7 V).

Ans. (a) $(10^5/16)(0.015 \pm j)$; (b) 1 kHz; (c) 21.36 V (peak-to-peak)

The circuit of Fig. 15.8 employs the resistance-variation mechanism of amplitude control, which we studied in the previous section (see Fig. 15.4). Potentiometer P is adjusted until oscillations just start to grow. As the oscillations grow, the diodes start to conduct, causing the effective resistance between a and b to decrease. Equilibrium will be reached at the output amplitude that causes the loop gain to be exactly one. The output amplitude can be varied by adjusting potentiometer P.

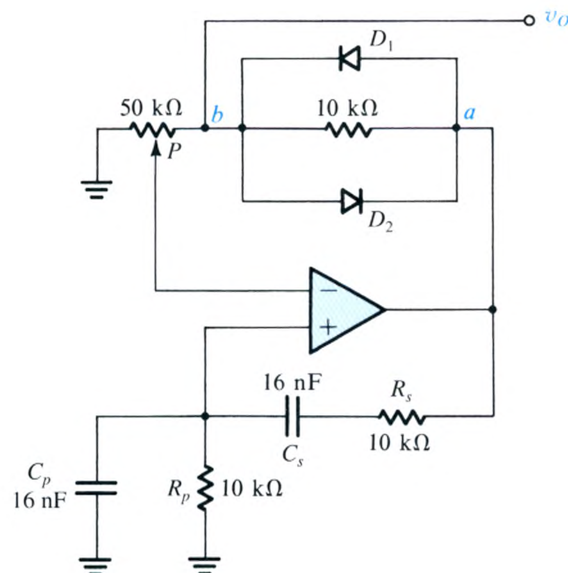


Figure 15.8 A Wien-bridge oscillator with an alternative method for amplitude stabilization.

As indicated in Fig. 15.8, the output is taken at point b rather than at the op-amp output terminal because the signal at b has lower distortion than that at a. To appreciate this point, note that the voltage at b is proportional to the voltage at the op-amp input terminals and that the latter is a filtered (by the RC network) version of the voltage at node a. Node b, however, is a high-impedance node, and a buffer will be needed if a load is to be connected.

EXERCISE

15.6 For the circuit in Fig. 15.8, find: (a) the setting of potentiometer P at which oscillations just start; (b) the frequency of oscillation.
Ans. (a) 20 kΩ to ground; (b) 1 kHz

15.2.2 The Phase-Shift Oscillator

The basic structure of the phase-shift oscillator is shown in Fig. 15.9. It consists of a negative-gain amplifier ($-K$) with a three-section (third-order) RC ladder network in the feedback. The circuit will oscillate at the frequency for which the phase shift of the RC network is 180°. Only at this frequency will the total phase shift around the loop be 0° or 360°. Here we should note that the reason for using a three-section RC network is that three is the minimum number of sections (i.e., lowest order) that is capable of producing a 180° phase shift at a finite frequency.

For oscillations to be sustained, the value of K should be equal to the inverse of the magnitude of the RC network transfer function at the frequency of oscillation. However, to ensure that oscillations start, the value of K has to be chosen slightly higher than the value that satisfies the unity-loop-gain condition. Oscillations will then grow in magnitude until limited by some nonlinear control mechanism.

Figure 15.10 shows a practical phase-shift oscillator with a feedback limiter, consisting of diodes D_1 and D_2 and resistors R_1, R_2, R_3, and R_4 for amplitude stabilization. To start oscillations, R_f has to be made slightly greater than the minimum required value. Although the circuit stabilizes more rapidly and provides sine waves with more stable amplitude if R_f is made much larger than this minimum, the price paid is an increased output distortion.

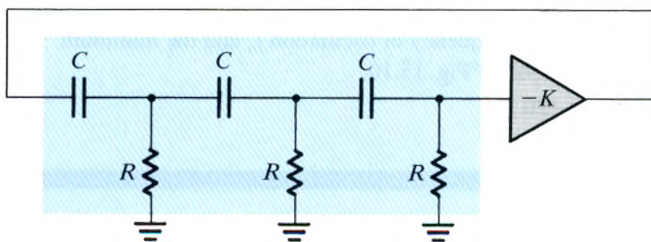


Figure 15.9 A phase-shift oscillator.

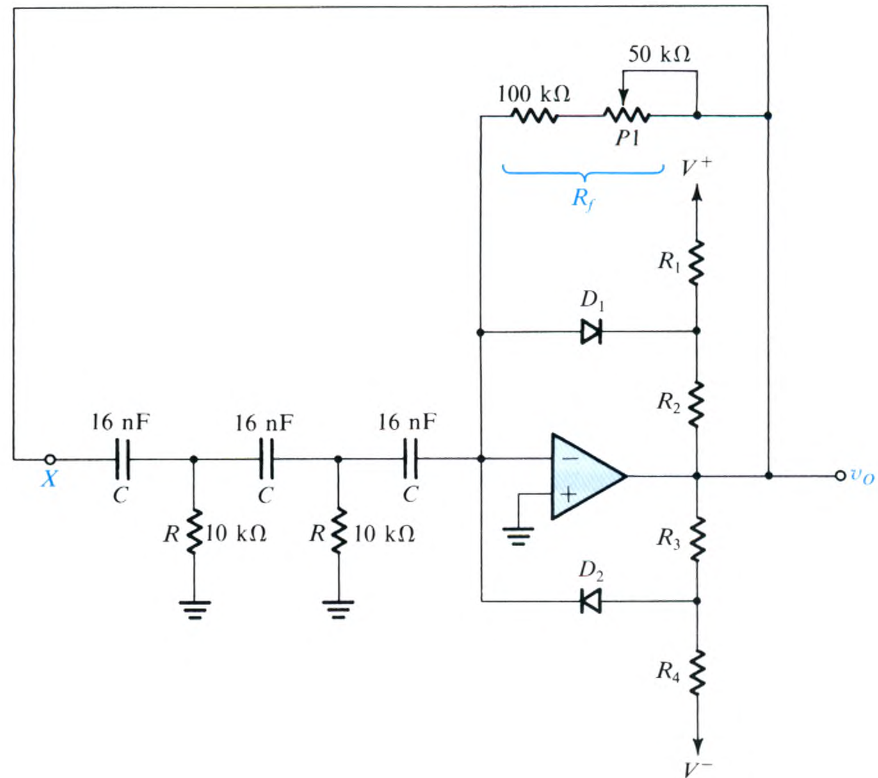


Figure 15.10 A practical phase-shift oscillator with a limiter for amplitude stabilization.

EXERCISES

15.7 Consider the circuit of Fig. 15.10 *without* the limiter. Break the feedback loop at X and find the loop gain $A\beta \equiv V_o(j\omega)/V_x(j\omega)$ in symbolic form (i.e., do not substitute the numerical values given). To do this, it is easier to start at the output and work backward, finding the various currents and voltages, and eventually V_x in terms of V_o.

Ans. $\dfrac{\omega^2 C^2 R R_f}{4 + j(3\omega CR - 1/\omega CR)}$

15.8 Use the expression derived in Exercise 15.7 to find the frequency of oscillation f_0 and the minimum required value of R_f for oscillations to start in the circuit of Fig. 15.10.

Ans. $\omega_0 = 1/\sqrt{3}CR; R_f \geq 12R; f_0 = 574.3\,\text{Hz}; R_f = 120\text{ k}\Omega$

15.2.3 The Quadrature Oscillator

The **quadrature oscillator** is based on the two-integrator loop studied in Section 14.6. As an active filter, the loop is damped to locate the poles in the left half of the s plane. Here, we will not use damping, since we want to locate the poles on the $j\omega$ axis to provide sustained oscillations. In fact, to ensure that oscillations start, the poles are initially located in the right half-plane and then "pulled back" by the nonlinear gain control.

Figure 15.11(a) shows a practical quadrature oscillator. Amplifier 1 is connected as an inverting Miller integrator with a limiter in the feedback for amplitude control. Amplifier 2 is connected as a noninverting integrator [thus replacing the cascade connection of the Miller integrator and the inverter in the two-integrator loop of Fig. 14.25(b)]. To understand the operation of this noninverting integrator, consider the equivalent circuit shown in Fig. 15.11(b). Here, we have replaced the integrator input voltage v_{O1} and the series resistance $2R$ by the Norton equivalent composed of a current source $v_{O1}/2R$ and a parallel resistance $2R$. Now, since $v_{O2} = 2v$, where v is the voltage at the input of op amp 2, the current through R_f will be $(2v - v)/R_f = v/R_f$ in the direction from output to input. Thus R_f gives rise to a negative input resistance, $-R_f$, as indicated in the equivalent circuit of Fig. 15.11(b). Nominally, R_f is made equal to $2R$, and thus $-R_f$ cancels $2R$, and at the input we are left with a current source $v_{O1}/2R$ feeding a capacitor C. The result is that $v = \frac{1}{C}\int_0^t \frac{v_{O1}}{2R}dt$ and $v_{O2} = 2v = \frac{1}{CR}\int_0^t v_{O1}dt$. That is, for $R_f = 2R$, the circuit functions as a perfect noninverting integrator. If, however, R_f is made smaller than $2R$, a net negative resistance appears in parallel with C.

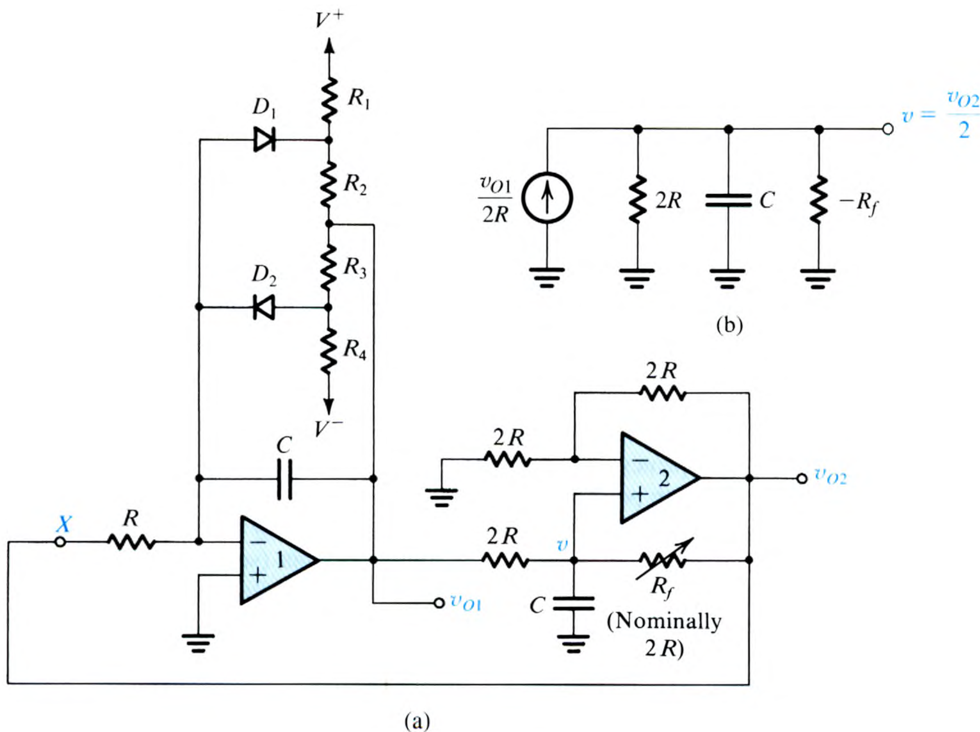


Figure 15.11 (a) A quadrature-oscillator circuit. (b) Equivalent circuit at the input of op amp 2.

Returning to the oscillator circuit in Fig. 15.11(a), we note that the resistance R_f in the positive-feedback path of op amp 2 is made variable, with a nominal value of $2R$. Decreasing the value of R_f moves the poles to the right half-plane (Problem 15.24) and ensures that the oscillations start. Too much positive feedback, although it results in better amplitude stability, also results in higher output distortion (because the limiter has to operate "harder"). In this regard, note that the output v_{O2} will be "purer" than v_{O1} because of the filtering action provided by the second integrator on the peak-limited output of the first integrator.

If we disregard the limiter and break the loop at X, we can find the loop gain as

$$L(s) \equiv \frac{V_{o2}}{V_x} = -\frac{1}{s^2C^2R^2} = \frac{1}{\omega^2C^2R^2} \tag{15.13}$$

Thus the loop will oscillate at frequency ω_0, given by

$$\omega_0 = \frac{1}{CR} \tag{15.14}$$

Finally, note that the name *quadrature oscillator* is used because the circuit provides two sinusoids with 90° phase difference. This is because v_{O2} is the integral of v_{O1}. There are many applications requiring quadrature sinusoids.

15.2.4 The Active-Filter-Tuned Oscillator

The last oscillator circuit that we will discuss is quite simple both in principle and in design. Nevertheless, the approach is general and versatile and can result in high-quality (i.e., low-distortion) output sine waves. The basic principle is illustrated in Fig. 15.12. The circuit consists of a high-Q bandpass filter connected in a positive-feedback loop with a hard limiter. To understand how this circuit works, assume that oscillations have already started. The output of the bandpass filter will be a sine wave whose frequency is equal to the center frequency of the filter, f_0. The sine-wave signal v_1 is fed to the limiter, which produces at its output a square wave whose levels are determined by the limiting levels and whose frequency is f_0. The square wave in turn is fed to the bandpass filter, which filters out the harmonics and provides a sinusoidal output v_1 at the fundamental frequency f_0. Obviously, the purity of the output sine wave will be a direct function of the selectivity (or Q factor) of the bandpass filter.

The simplicity of this approach to oscillator design should be apparent. We have independent control of frequency and amplitude as well as of distortion of the output sinusoid. Any filter circuit with positive gain can be used to implement the bandpass filter. The frequency stability of the oscillator will be directly determined by the frequency stability of the bandpass-filter circuit. Also, a variety of limiter circuits (see Section 4.1.3) with different degrees of sophistication can be used to implement the limiter block.

Figure 15.13 shows one possible implementation of the active-filter-tuned oscillator. This circuit uses a variation on the bandpass circuit based on the Antoniou inductance-simulation circuit [see Fig. 14.22(b)]. Here resistor R_2 and capacitor C_4 are interchanged. This makes the output of the lower op amp directly proportional to (in fact, twice as large as) the voltage across the resonator, and we can therefore dispense with the buffer amplifier K. The limiter used is a very simple one consisting of a resistance R_1 and two diodes.

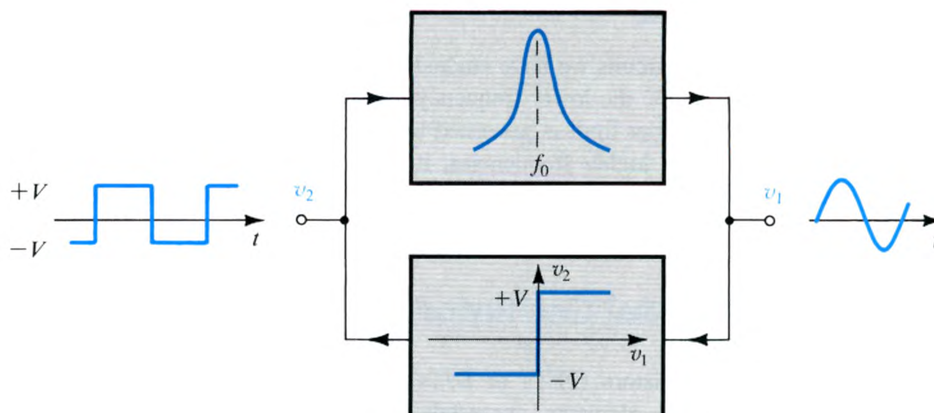


Figure 15.12 Block diagram of the active-filter-tuned oscillator.

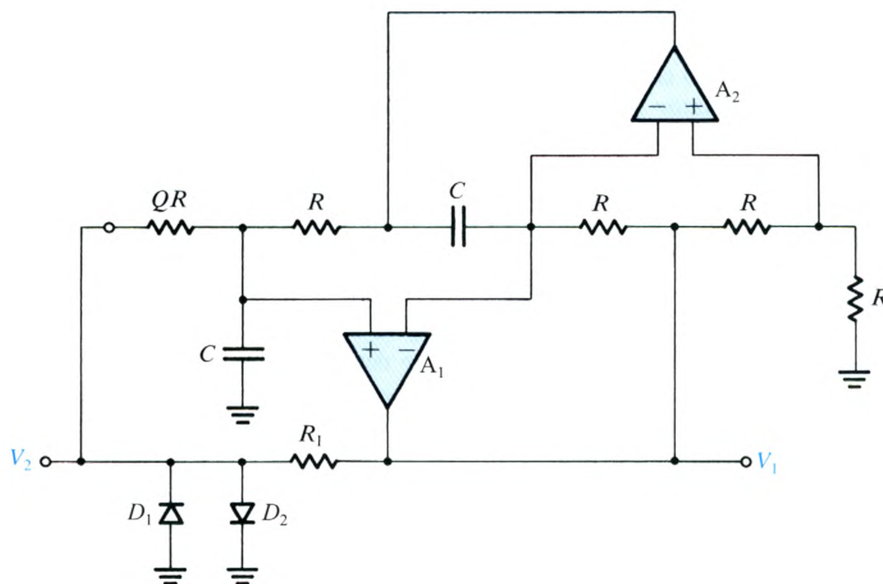


Figure 15.13 A practical implementation of the active-filter-tuned oscillator.

EXERCISE

15.9 Using $C = 16\,\text{nF}$, find the value of R required for the circuit in Fig. 15.13 to produce 1-kHz sine waves. If the diode drop is 0.7 V, find the peak-to-peak amplitude of the output sine wave. (*Hint:* A square wave with peak-to-peak amplitude of V volts has a fundamental component with $4V/\pi$ volts peak-to-peak amplitude.)

Ans. 10 kΩ; 3.6 V

15.2.5 A Final Remark

The op amp–RC oscillator circuits we have studied are useful when operating in the range 10 Hz to 1 MHz or so. While the lower frequency limit is dictated by the size of passive components required, the upper limit is governed by the frequency-response and slew-rate limitations of op amps. For higher frequencies, it is common to use circuits employing transistors together with LC-tuned circuits or crystals.[1] These are discussed in Section 15.3.

15.3 LC and Crystal Oscillators

Oscillators employing transistors (FETs or BJTs), with LC circuits or crystals as the frequency-selective feedback elements, are used in the frequency range of 100 kHz to hundreds of gigahertz. They exhibit higher Q than the RC types. However, LC oscillators are difficult to tune over wide ranges, and crystal oscillators operate at a single frequency.

15.3.1 The Colpitts and Hartely Oscillators

Figure 15.14 shows two commonly used configurations of LC oscillators. They are known as the **Colpitts oscillator** and the **Hartley oscillator**. Both utilize a parallel LC circuit connected between the drain and the gate of a MOSFET, with a fraction of the tuned-circuit voltage applied between the gate and the source. This feedback is achieved by way of a capacitive divider in the Colpitts oscillator and by way of an inductive divider in the Hartley circuit. Observe that in both circuits the voltage V_{sg} gives rise to a current I in the direction shown, which in turn results in a positive voltage across the LC circuit and thus in a positive V_{sg}. Thus, we do have a positive-feedback loop.

If the frequency of operation is sufficiently low that we can neglect the transistor capacitances, the frequency of oscillation will be determined by the resonance frequency of the parallel-tuned circuit (also known as a *tank circuit* because it behaves as a reservoir

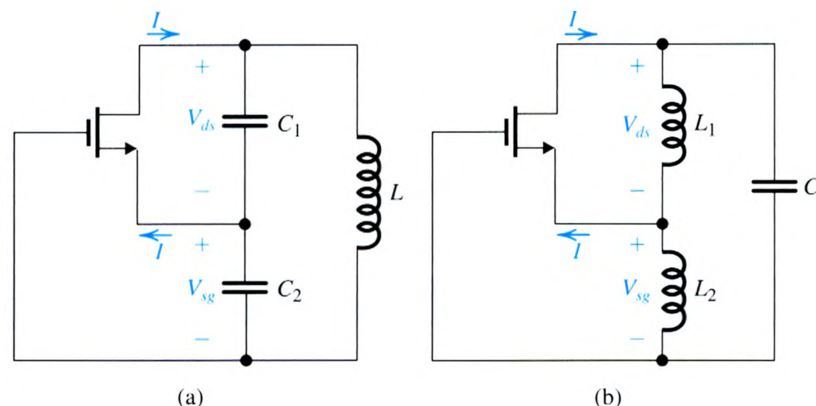


Figure 15.14 Two commonly used configurations of LC-tuned oscillators: **(a)** Colpitts and **(b)** Hartley.

[1] Of course, transistors can be used in place of the op amps in the circuits just studied. At higher frequencies, however, better results are obtained with LC-tuned circuits and crystals.

for energy storage). Thus for the Colpitts oscillator we have

$$\omega_0 = 1 \Bigg/ \sqrt{L\left(\frac{C_1 C_2}{C_1 + C_2}\right)} \tag{15.15}$$

and for the Hartley oscillator we have

$$\omega_0 = 1/\sqrt{(L_1 + L_2)C} \tag{15.16}$$

The ratio L_1/L_2 or C_1/C_2 determines the feedback factor and thus must be adjusted in conjunction with the transistor gain to ensure that oscillations will start. Specifically, the output voltage of the MOSFET, V_{ds}, is related to the input voltage V_{sg} by

$$V_{ds} = A\, V_{sg} \tag{15.17}$$

where A, the voltage gain of the MOSFET, is given by

$$A = g_m R \tag{15.18}$$

where R is the total effective resistance between drain and source. Resistance R includes r_o of the MOSFET as well as a resistance that represents the inductor loss (i.e., represents the fact the LC circuit has a finite Q factor).

Now, at $\omega = \omega_0$, the LC circuit has infinite impedance, which makes the current $I = 0$. Thus, the voltage V_{ds} across C_1 (in the Colpitts oscillator) gives rise to a current sC_1V_{ds} that flows through the series combination of C_1 and C_2. This current in turn gives rise to a voltage V_{sg} across C_2 given by

$$V_{sg} = sC_1V_{ds} \times \frac{1}{sC_2} = \frac{C_1}{C_2} V_{ds}$$

Thus, the feedback factor is given by

$$\beta = \frac{V_{sg}}{V_{ds}} = \frac{C_1}{C_2} \tag{15.19}$$

and using Eqs. (15.17) and (15.18), the loop gain is found as

$$A\beta = A\,\frac{C_1}{C_2} = (g_m R)\left(\frac{C_1}{C_2}\right) \tag{15.20}$$

For oscillations to start, we make

$$(g_m R)\left(\frac{C_1}{C_2}\right) > 1 \tag{15.21}$$

or equivalently

$$g_m R > \frac{C_2}{C_1} \tag{15.22}$$

As oscillations grow in amplitude, the transistor's nonlinear characteristics reduce the effective value of g_m and, correspondingly, reduce the loop gain to unity, thus

$$g_m R = \frac{C_2}{C_1} \tag{15.23}$$

which is the condition for sustained oscillations.

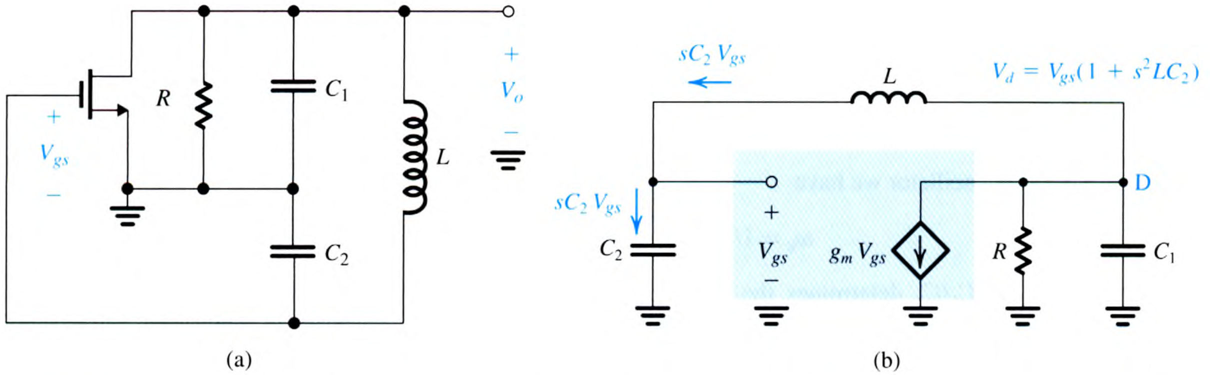


Figure 15.15 (**a**) A Colpitts oscillator in which the source is grounded and the output is taken at the drain. (**b**) Equivalent circuit of the Colpitts oscillator of (**a**). To simplify the analysis, ϵ_{gd} is neglected. We can consider C_{gs} to be part of C_2, and we can include r_o in R.

The condition of oscillation for the Hartley oscillator in Fig. 15.14(b) can be similarly derived with the result that

$$g_mR > \frac{L_1}{L_2} \tag{15.24}$$

for oscillations to start and

$$g_mR = \frac{L_1}{L_2} \tag{15.25}$$

for sustained oscillations.

Depending on where the output voltage of the oscillator is to be taken, the appropriate terminal of the transistor can be connected to ground. In this regard, note that any of the three terminals can be grounded without changing the nature of the feedback loop. As an example, we show in Fig. 15.15(a) the Colpitts oscillator with the emitter connected to ground and the output taken at the collector. Although the bias arrangement is not shown, we have included a resistance R that models the combination of the output resistance of the transistor (r_o), the inductor loss, and the input resistance of the circuit to which the oscillator output is connected. Other circuit configurations with complete bias detail are explored in the end-of-chapter problems.

As an alternative to the analysis above, we can determine the conditions of oscillation of the Colpitts oscillator in Fig. 15.15(a) as follows: We replace the transistor with its equivalent circuit, as shown in Fig. 15.15(b). To simplify the analysis, we have neglected the transistor capacitance C_{gd}. Capacitance C_{gs}, although not shown, can be considered to be a part of C_2.

Assuming oscillations already exist, we analyze the circuit as shown in Fig. 15.15(b). A node equation at the transistor drain (node D) yields

$$sC_2V_{gs} + g_mV_{gs} + \left(\frac{1}{R} + sC_1\right)(1 + s^2LC_2)V_{gs} = 0$$

Since $V_{gs} \neq 0$ (oscillations have started), it can be eliminated, and the equation can be rearranged in the form

$$s^3LC_1C_2 + s^2(LC_2/R) + s(C_1 + C_2) + \left(g_m + \frac{1}{R}\right) = 0 \tag{15.26}$$

Substituting $s = j\omega$ gives

$$\left(g_m + \frac{1}{R} - \frac{\omega^2 LC_2}{R}\right) + j[\omega(C_1 + C_2) - \omega^3 LC_1C_2] = 0 \tag{15.27}$$

For oscillations to start, both the real and imaginary parts must be zero. Equating the imaginary part to zero gives the frequency of oscillation as

$$\omega_0 = 1 \Bigg/ \sqrt{L\left(\frac{C_1C_2}{C_1 + C_2}\right)} \tag{15.28}$$

which is the resonance frequency of the tank circuit, as anticipated. Equating the real part to zero and using Eq. (15.28) gives

$$g_m R = C_2/C_1 \tag{15.29}$$

which is the condition for sustained oscillations.

We conclude with a few words on the mechanism that determines the amplitude of oscillations in the LC oscillators discussed above. Unlike the op-amp oscillators that incorporate special amplitude-control circuitry, LC oscillators utilize the nonlinear i_D–v_{GS} characteristics of the MOSFET for amplitude control. Thus these LC oscillators are known as *self-limiting oscillators*. Specifically, as the oscillations grow in amplitude, the effective gain of the transistor is reduced below its small-signal value. Eventually, they reach an amplitude at which the effective gain is reduced to the point that the Barkhausen criterion is satisfied exactly. The amplitude then remains constant at this value.

Reliance on the nonlinear characteristics of the MOSFET implies that the drain current waveform will be nonlinearly distorted. Nevertheless, the output voltage signal will still be a sinusoid of high purity because of the filtering action of the LC circuit.

EXERCISES

15.10 For the Colpitts oscillator in Fig 15.14(a), show that the admittance Y of the tuned circuit seen by the transistor between the drain and the source is zero at $\omega = \omega_0$. (This is the reason the current $I = 0$.)

D15.11 Using a MOSFET biased to operate at $g_m = 5$ mA/V, design a Colpitts oscillator to operate at $\omega_0 = 10^6$ rad/s. Use $C_1 = 0.01\ \mu$F and assume that the coil available has a Q of 100 (this can be represented by a resistance in parallel with C_1 given by $Q/\omega_0 C_1$). Also assume that there is a load resistance at the collector of 100 kΩ and that for the MOSFET, $r_o = 100$ kΩ. Find C_2 and L.

Ans. 0.42 μF; 102 μH (a somewhat smaller C_2 would be used to allow oscillations to grow in amplitude)

OSCILLATOR PIONEERS

Heinrich Georg Barkhausen was a German physicist who taught at the Technical University of Dresden. In 1911, at age 29, he became what has since been described as the world's first chaired professor of electrical engineering. Among his many interests were high-frequency oscillators, and he created the enduring Barkhausen stability criterion.

Nevada-born **Ralph Hartley** was a Rhodes Scholar who began work at Western Electric in 1915. He led a transatlantic-radio project for which he invented his tapped-inductor oscillator. As well, he is renowned for having developed a neutralizing scheme to prevent the triode "singing" (self-oscillation) produced by internal input/output capacitance coupling (analogous to that produced by C_μ in a BJT or C_{gd} in a FET).

Edwin Henry Colpitts was a school principal in Newfoundland. He later obtained advanced education in math and physics at Harvard and in 1907 joined Western Electric, where he ultimately led a group in radio telephony. Later, he improved upon the Hartley tapped-inductor oscillator by introducing a more flexible capacitance voltage divider. The circuit now known as the Colpitts oscillator was patented in 1920 as the "oscillation generator." Though invented in the context of vacuum tubes, the Colpitts oscillator is extremely versatile and has marked significance in BJT and MOS designs.

15.3.2 The Cross-Coupled LC Oscillator

Figure 15.16(a) shows a popular LC oscillator circuit suitable for fabrication in IC form and capable of operating at frequencies approaching hundreds of gigahertz (for use in wireless transceivers). It consists of a pair of MOSFETs connected in the differential amplifier configuration, each with a parallel LC-tuned circuit load, and with the drain of each connected to the gate of the other. The latter connection gives rise to the "cross-coupled" part of the name.

To see how the cross-coupled LC oscillator works, first note that, as in a differential amplifier, from a signal point of view each of Q_1 and Q_2 operates in the grounded-source configuration. Next, observe that if we think of Q_1 and Q_2 together with their loads as common-source amplifiers, we see that the cross coupling simply means that the output of each amplifier drives the input of the other, resulting in the feedback loop shown in Fig. 15.16(b). Here we have eliminated all dc bias sources to concentrate on the signal operation of the circuit.

Examination of the feedback loop in Fig. 15.16(b) reveals that at the resonance frequency of each of the two tank circuits (i.e., at $\omega = \omega_0 = 1\sqrt{LC}$), the load of each of Q_1 and Q_2 reduces to a resistance $R_p = \omega_0 LQ$, where Q is the quality factor of the inductance. Taking into consideration the output resistance r_o of each of Q_1 and Q_2, we can write for the gain of each of the two stages at $\omega = \omega_0$,

$$A_1 = A_2 = -g_m(R_p \parallel r_o)$$

Thus each stage exhibits a 180° phase shift, for a total phase shift around the loop of 360°. It follows that the circuit will provide sustained oscillations at

$$\omega_0 = 1\sqrt{LC} \tag{15.30}$$

provided

$$|A_1A_2| = [g_m(R_p \parallel r_o)]^2 = 1$$

which reduces to

$$g_m(R_p \parallel r_o) = 1 \tag{15.31}$$

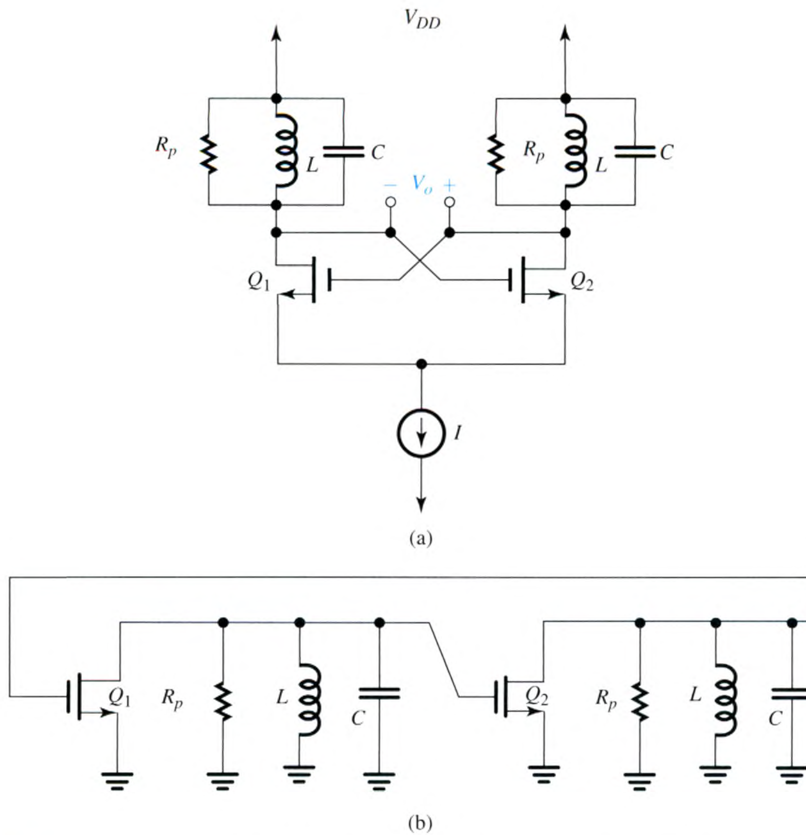


Figure 15.16 **(a)** The cross-coupled LC oscillator. **(b)** Signal equivalent circuit of the cross-coupled oscillator in **(a)**.

The condition in Eq. (15.31) can be used to determine the minimum required value of g_m at which each of Q_1 and Q_2 is operated for oscillations to be sustained. As usual, to ensure that oscillations start, a somewhat higher value of g_m is used. Amplitude stabilization is provided by the nonlinear MOSFET characteristics.

For applications at very high frequency, the inductor is fabricated on chip by depositing a thin metal film in a spiral shape. Such IC inductors have small values (in the nanohenry range) and, unfortunately, small Q factors as well.

EXERCISE

D15.12 Design the cross-coupled oscillator to operate at $\omega_0 = 10$ Grad/s. The IC inductors available have $L = 10$ nH and $Q = 10$. If the transistor $r_o = 10$ kΩ, find the required value of C and the minimum required value of g_m at which Q_1 and Q_2 are to be operated.
Ans. 1 pF; 1.1 mA/V

15.3.3 Crystal Oscillators

A piezoelectric crystal, such as quartz, exhibits electromechanical-resonance characteristics that are very stable (with time and temperature) and highly selective (having very high Q factors). The circuit symbol of a crystal is shown in Fig. 15.17(a), and its equivalent-circuit model is given in Fig. 15.17(b). The resonance properties are characterized by a large inductance L (as high as hundreds of henrys), a very small series capacitance C_s (as small as 0.0005 pF), a series resistance r representing a Q factor $\omega_0 L/r$ that can be as high as a few hundred thousand, and a parallel capacitance C_p (a few picofarads). Capacitor C_p represents the electrostatic capacitance between the two parallel plates of the crystal. Note that $C_p \gg C_s$.

Since the Q factor is very high, we can neglect the resistance r and express the crystal impedance as

$$Z(s) = 1 \bigg/ \left[sC_p + \frac{1}{sL + 1/sC_s} \right]$$

which can be manipulated to the form

$$Z(s) = \frac{1}{sC_p} \frac{s^2 + (1/LC_s)}{s^2 + \left[(C_p + C_s)/LC_sC_p\right]} \tag{15.32}$$

From Eq. (15.32) and from Fig. 15.17(b), we see that the crystal has two resonance frequencies: a series resonance at ω_s

$$\omega_s = 1/\sqrt{LC_s} \tag{15.33}$$

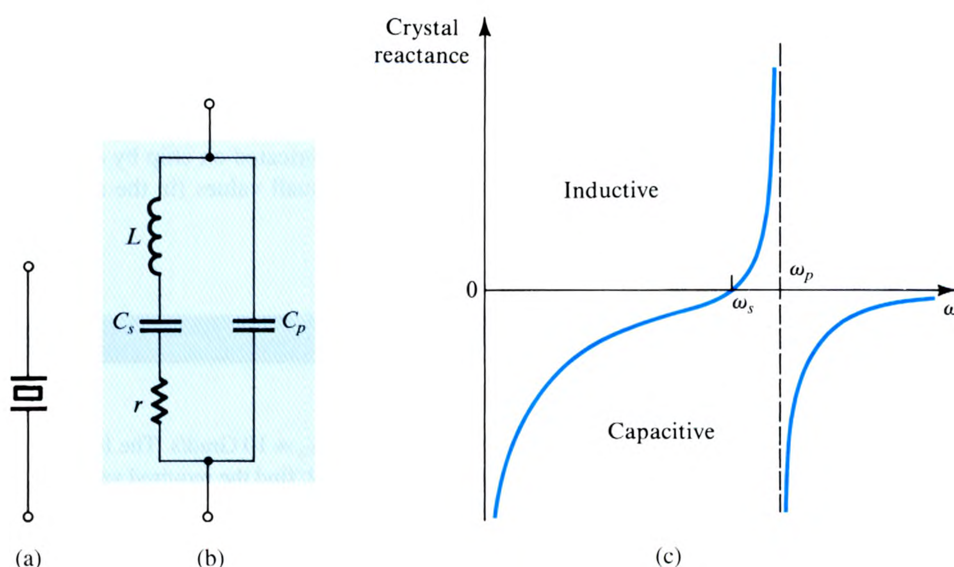


Figure 15.17 A piezoelectric crystal. **(a)** Circuit symbol. **(b)** Equivalent circuit. **(c)** Crystal reactance versus frequency [note that, neglecting the small resistance r, $Z_{\text{crystal}} = jX(\omega)$].

and a parallel resonance at ω_p

$$\omega_p = 1 \Bigg/ \sqrt{L\left(\frac{C_s C_p}{C_s + C_p}\right)} \tag{15.34}$$

Thus for $s = j\omega$ we can write

$$Z(j\omega) = -j\frac{1}{\omega C_p}\left(\frac{\omega^2 - \omega_s^2}{\omega^2 - \omega_p^2}\right) \tag{15.35}$$

From Eqs. (15.33) and (15.34) we note that $\omega_p > \omega_s$. However, since $C_p \gg C_s$, the two resonance frequencies are very close. Expressing $Z(j\omega) = jX(\omega)$, the crystal reactance $X(\omega)$ will have the shape shown in Fig. 15.17(c). We observe that the crystal reactance is inductive over the very narrow frequency band between ω_s and ω_p. For a given crystal, this frequency band is well defined. This means that we can use the crystal to replace the inductor of the Colpitts oscillator [Fig. 15.14(a)]. The resulting circuit will oscillate at the resonance frequency of the crystal inductance L with the series equivalent of C_s and $(C_p + C_1C_2/(C_1 + C_2))$. Since C_s is much smaller than the three other capacitances, it will be dominant and

$$\omega_0 \simeq 1/\sqrt{LC_s} = \omega_s \tag{15.36}$$

In addition to the basic Colpitts oscillator, there are various configurations for crystal oscillators. Figure 15.18 shows a popular configuration (called the **Pierce oscillator**) using a CMOS inverter (see Section 16.3) as an amplifier. Resistor R_f determines a dc operating point in the high-gain region of the VTC of the CMOS inverter. Resistor R_1 together with capacitor C_1 provides a low-pass filter that discourages the circuit from oscillating at a higher harmonic of the crystal frequency. Note that this circuit also is based on the Colpitts configuration.

The extremely stable resonance characteristics and the very high Q factors of quartz crystals result in oscillators with very accurate and stable frequencies. Crystals are available with resonance frequencies in the range of a few kilohertz to hundreds of megahertz.

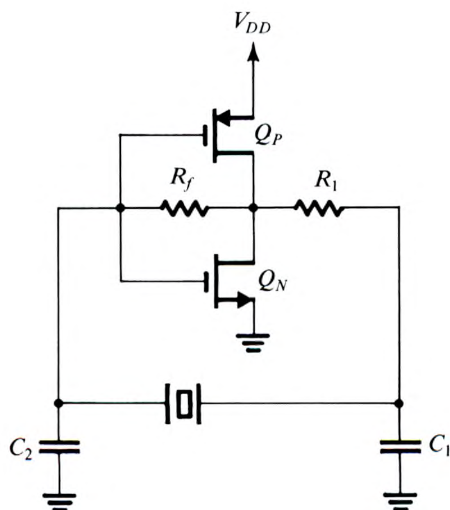


Figure 15.18 A Pierce crystal oscillator using a CMOS inverter as an amplifier.

Temperature coefficients of ω_0 of 1 or 2 parts per million (ppm) per degree Celsius are achievable. Unfortunately, however, crystal oscillators, being mechanical resonators, are fixed-frequency circuits.

EXERCISE

15.13 A 2-MHz quartz crystal is specified to have $L = 0.52$ H, $C_s = 0.012$ pF, $C_p = 4$ pF, and $r = 120\ \Omega$. Find f_s, f_p, and Q.
Ans. 2.015 MHz; 2.018.MHz; 55,000

15.4 Nonlinear Oscillators or Function Generators

The oscillators we have studied so far generate sine waves by connecting a frequency-selective network (a filter) in the feedback path of a positive gain amplifier. Although a nonlinear circuit is used for amplitude control, the conditions of oscillation are determined using linear circuit analysis, and these oscillators are correspondingly known as linear oscillators. By contrast, the oscillator circuits we will study in this section use nonlinear circuits to generate square and triangular waveforms. The basic building block of these nonlinear oscillators is the **bistable multivibrator**, which we will study first.

As its name indicates, the bistable multivibrator has *two stable states*. The circuit can remain in either stable state indefinitely and moves to the other stable state only when appropriately *triggered*.

15.4.1 The Bistable Feedback Loop

Bistability can be obtained by connecting a dc amplifier in a positive-feedback loop with a loop gain greater than one. Such a feedback loop is shown in Fig. 15.19; it consists of an op amp and a resistive voltage divider in the positive-feedback path. To see how bistability is obtained, consider operation with the positive input terminal of the op amp near ground potential. This is a reasonable starting point since the circuit has no external excitation. Assume that the electrical noise that is inevitably present in every electronic circuit causes a small positive increment in the voltage v_+. This incremental signal will be amplified by the large open-loop gain A of the op amp, with the result that a much greater positive increment will appear in the op amp's output, v_O. The voltage divider (R_1, R_2) will feed a fraction $\beta \equiv R_1/(R_1 + R_2)$ of the output signal back to the positive input terminal of the op amp. If $A\beta$ is greater than unity, as is usually the case, the fed-back signal will be greater than the original increment in v_+. This *regenerative* process continues until eventually the op amp saturates with its output voltage at the positive saturation level, L_+. When this happens, the voltage at the positive input terminal, v_+, becomes $L_+R_1/(R_1 + R_2)$, which is positive and thus keeps the op amp in positive saturation. This is one of the two stable states of the circuit.

In the description above we assumed that when v_+ was near zero volts, a positive increment occurred in v_+. Had we assumed the equally probable situation of a negative

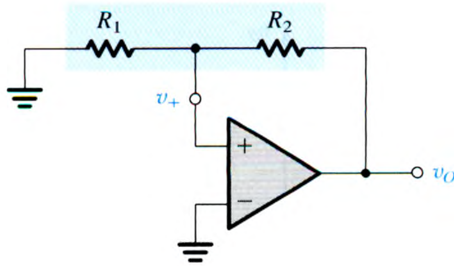


Figure 15.19 A positive-feedback loop capable of bistable operation.

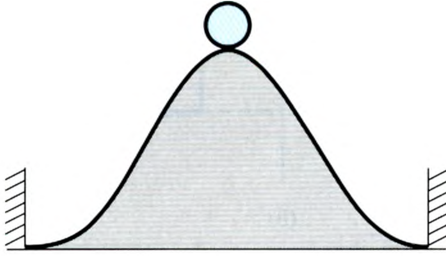


Figure 15.20 A physical analogy for the operation of the bistable circuit. The ball cannot remain at the top of the hill for any length of time (a state of unstable equilibrium or metastability); the inevitably present disturbance will cause the ball to fall to one side or the other, where it can remain indefinitely (the two stable states).

increment, the op amp would have ended up saturated in the negative direction with $v_O = L_-$ and $v_+ = L_- R_1/(R_1 + R_2)$. This is the other stable state.

We thus conclude that the circuit of Fig. 15.19 has two stable states, one with the op amp in positive saturation and the other with the op amp in negative saturation. The circuit can exist in either of these two states indefinitely. We also note that the circuit cannot exist in the state for which $v_+ = 0$ and $v_O = 0$ for any length of time. This is a state of *unstable equilibrium* (also known as a **metastable state**); any disturbance, such as that caused by electrical noise, causes the bistable circuit to switch to one of its two stable states. This is in sharp contrast to the case when the feedback is negative, causing a virtual short circuit to appear between the op amp's input terminals and maintaining this virtual short circuit in the face of disturbances. A physical analogy for the operation of the bistable circuit is depicted in Fig. 15.20.

15.4.2 Transfer Characteristic of the Bistable Circuit

You may be wondering how we can make the bistable circuit of Fig. 15.19 change state. Deriving the transfer characteristic of the bistable circuit will help to answer this question. Figure 15.19 shows that there are two nodes connected to ground. Either of these nodes can serve as an input terminal. Let us investigate both possibiliies.

Figure 15.21(a) shows the bistable circuit with a voltage v_I applied to the inverting input terminal of the op amp. To derive the transfer characteristic v_O–v_I, assume that v_O is at one of its two possible levels, say L_+, and thus $v_+ = \beta L_+$. Now as v_I is increased from 0 V, we can see from the circuit that nothing happens until v_I reaches a value equal to v_+ (i.e., βL_+). As v_I begins to exceed this value, a net negative voltage develops between the input terminals of the op amp. This voltage is amplified by the open-loop gain of the op amp, and thus v_O goes negative. The voltage divider in turn causes v_+ to go negative, thus increasing the net negative input to the op amp and keeping the regenerative process going. This process culminates in the op amp saturating in the negative direction: that is, with $v_O = L_-$ and, correspondingly, $v_+ = \beta L_-$. It is easy to see that increasing v_I further has no effect on the acquired state of the bistable circuit. Figure 15.21(b) shows the transfer characteristic for increasing v_I. Notice

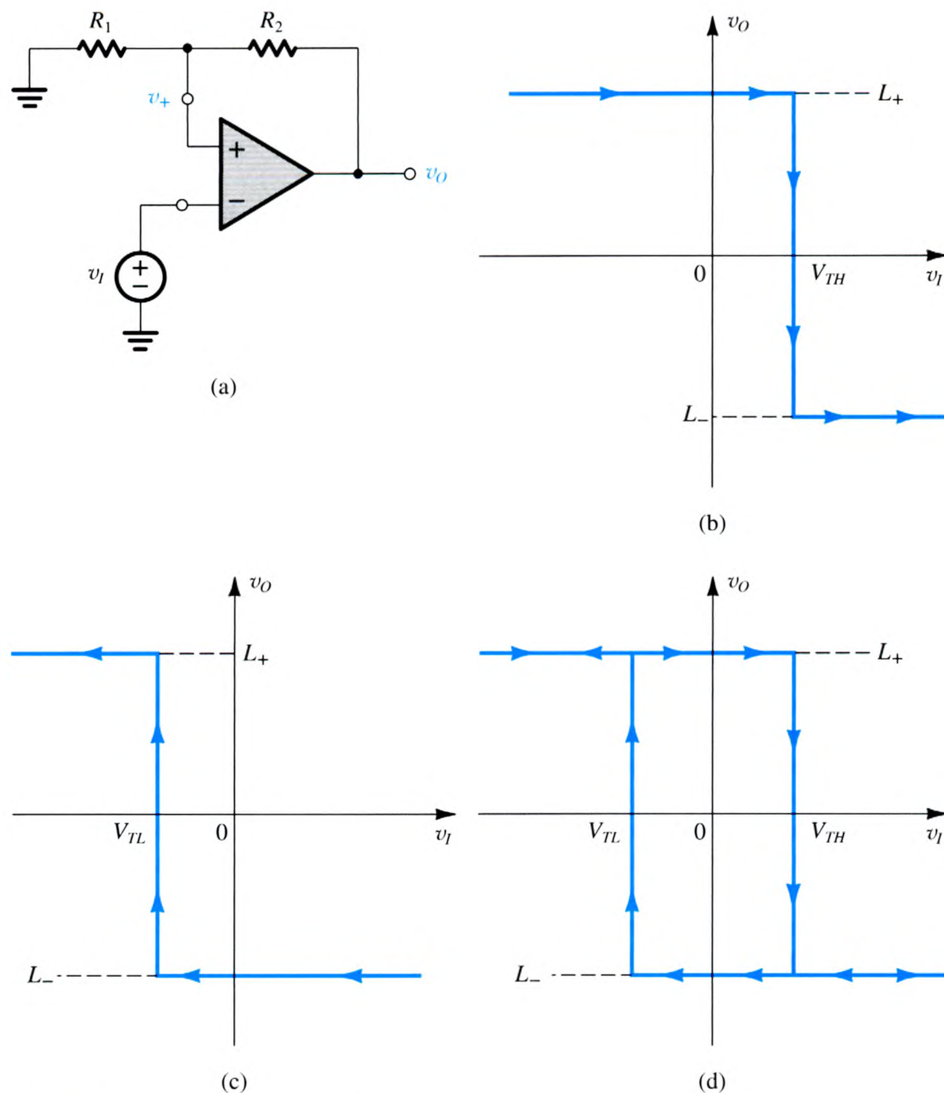


Figure 15.21 **(a)** The bistable circuit of Fig. 15.19 with the negative input terminal of the op amp disconnected from ground and connected to an input signal v_I. **(b)** The transfer characteristic of the circuit in **(a)** for increasing v_I. **(c)** The transfer characteristic for decreasing v_I. **(d)** The complete transfer characteristics.

that the characteristic is that of a comparator with a threshold voltage denoted V_{TH}, where $V_{TH} = \beta L_+$.

Now consider what happens as v_I is decreased. Since now $v_+ = \beta L_-$, we see that the circuit remains in the negative-saturation state until v_I goes negative to the point that it equals βL_-. As v_I goes below this value, a net positive voltage appears between the op amp's input terminals. This voltage is amplified by the op-amp gain and thus gives rise to a positive voltage at the op amp's output. The regenerative action of the positive-feedback loop then sets in and causes the circuit eventually to go to its positive saturation state, in which $v_O = L_+$ and $v_+ = \beta L_+$. The transfer characteristic for decreasing v_I is shown in Fig. 15.21(c). Here

again we observe that the characteristic is that of a comparator, but with a threshold voltage $V_{TL} = \beta L_-$.

The complete transfer characteristics, v_O–v_I, of the circuit in Fig. 15.21(a) can be obtained by combining the characteristics in Fig. 15.21(b) and (c), as shown in Fig. 15.21(d). As indicated, the circuit changes state at different values of v_I, depending on whether v_I is increasing or decreasing. Thus the circuit is said to exhibit *hysteresis;* the width of the hysteresis is the difference between the high threshold V_{TH} and the low threshold V_{TL}. Also note that the bistable circuit is in effect a comparator with hysteresis. Finally, notice that because the bistable circuit of Fig. 15.21 switches from the positive state ($v_O = L_+$) to the negative state ($v_O = L_-$) as v_I is increased past the positive threshold V_{TH}, the circuit is said to be *inverting*. A bistable circuit with a *noninverting* transfer characteristic will be presented shortly.

15.4.3 Triggering the Bistable Circuit

Returning now to the question of how to make the bistable circuit change state, we observe from the transfer characteristics of Fig. 15.21(d) that if the circuit is in the L_+ state it can be switched to the L_- state by applying an input v_I of value greater than $V_{TH} \equiv \beta L_+$. Such an input causes a net negative voltage to appear between the input terminals of the op amp, which initiates the regenerative cycle that culminates in the circuit switching to the L_- stable state. Here it is important to note that the input v_I merely initiates or *triggers* regeneration. Thus we can remove v_I after regeneration has started with no effect on the completion of the regeneration process. In other words, v_I can be simply a pulse of short duration. The input signal v_I is thus referred to as a **trigger signal**, or simply a **trigger**.

The characteristics of Fig. 15.21(d) also indicate that the bistable circuit can be switched to the positive state ($v_O = L_+$) by applying a negative trigger signal v_I of magnitude greater than that of the negative threshold V_{TL}.

15.4.4 The Bistable Circuit as a Memory Element

We can see from Fig. 15.21(d) that for input voltages in the range $V_{TL} < v_I < V_{TH}$, the output can be either L_+ or L_-, *depending on the state that the circuit is already in*. So, for this input range, the output is determined by the *previous* value of the trigger signal (the trigger signal that caused the circuit to be in its current state). Thus the circuit exhibits *memory*. Indeed, the bistable multivibrator is the basic memory element of digital systems,[2] as we will see in Chapter 18. Finally, note that in analog-circuit applications, such as the ones of concern to us in this chapter, the bistable circuit is also known as a **Schmitt trigger**.

15.4.5 A Bistable Circuit with Noninverting Transfer Characteristic

The basic bistable feedback loop of Fig. 15.19 can be used to derive a circuit with noninverting transfer characteristic by applying the input signal v_I (the trigger signal) to the terminal of R_1 that is connected to ground. The resulting circuit is shown in Fig. 15.22(a). To obtain the transfer characteristic, we first apply superposition to the linear circuit formed by

[2]Digital implementations of multivibrators (Chapter 18) use MOSFETs. The op-amp circuits presented here are intended for precision analog systems.

R_1 and R_2, thus expressing v_+ in terms of v_I and v_O as

$$v_+ = v_I \frac{R_2}{R_1 + R_2} + v_O \frac{R_1}{R_1 + R_2} \tag{15.37}$$

From this equation we see that if the circuit is in the positive stable state with $v_O = L_+$, positive values for v_I will have no effect. To trigger the circuit into the L_- state, v_I must be made negative and of such a value as to make v_+ decrease below zero. Thus the low-threshold V_{TL} can be found by substituting in Eq. (15.37) $v_O = L_+$, $v_+ = 0$, and $v_I = V_{TL}$. The result is

$$V_{TL} = -L_+(R_1/R_2) \tag{15.38}$$

Similarly, Eq. (15.37) indicates that when the circuit is in the negative-output state ($v_O = L_-$), negative values of v_I will make v_+ more negative with no effect on operation. To initiate the regeneration process that causes the circuit to switch to the positive state, v_+ must be made to go slightly positive. The value of v_I that causes this to happen is the high-threshold voltage V_{TH}, which can be found by substituting in Eq. (15.37) $v_O = L_-$ and $v_+ = 0$. The result is

$$V_{TH} = -L_-(R_1/R_2) \tag{15.39}$$

The complete transfer characteristic of the circuit of Fig. 15.22(a) is displayed in Fig. 15.22(b). Observe that a positive triggering signal v_I (of value greater than V_{TH}) causes the circuit to switch to the positive state (v_O goes from L_- to L_+). Thus the transfer characteristic of this circuit is noninverting.

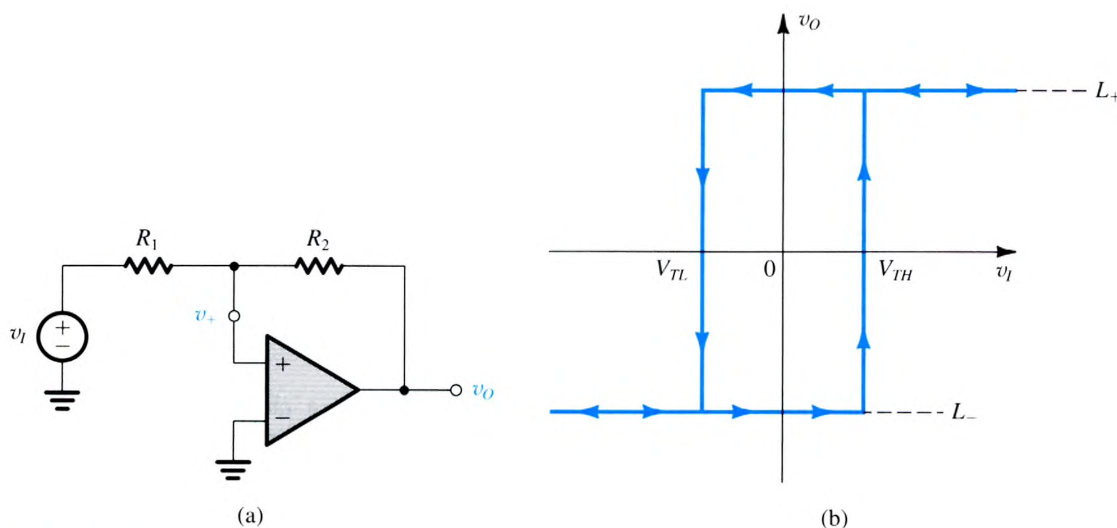


Figure 15.22 **(a)** A bistable circuit derived from the positive-feedback loop of Fig. 15.19 by applying v_I through R_1. **(b)** The transfer characteristic of the circuit in **(a)** is noninverting. [Compare it to the inverting characteristic in Fig. 15.21(d).]

EXERCISES

D15.14 The op amp in the bistable circuit of Fig. 15.21(a) has output saturation voltages of ±13 V. Design the circuit to obtain threshold voltages of ±5 V. For $R_1 = 10$ kΩ, find the value required for R_2.
Ans. 16 kΩ

D15.15 If the op amp in the circuit of Fig. 15.22(a) has ±10-V output saturation levels, design the circuit to obtain ±5-V thresholds. Give suitable component values.
Ans. Possible choice: $R_1 = 10$ kΩ and $R_2 = 20$ kΩ

15.16 Consider a bistable circuit with a noninverting transfer characteristic and let $L_+ = -L_- = 10$ V and $V_{TH} = -V_{TL} = 5$ V. If v_I is a triangular wave with a 0-V average, a 10-V peak amplitude, and a 1-ms period, sketch the waveform of v_O. Find the time interval between the zero crossings of v_I and v_O.
Ans. v_O is a square wave with 0-V average, 10-V amplitude, and 1-ms period and is delayed by 125 μs relative to v_I

15.4.6 Generating Square Waveforms Using a Bistable Circuit

We can generate a square waveform by arranging for a bistable circuit to switch states periodically. This can be done by connecting the bistable with an RC circuit in a feedback loop, as shown in Fig. 15.23(a). Notice that the bistable has an inverting transfer characteristic and can thus be realized using the circuit of Fig. 15.21(a). This results in the circuit

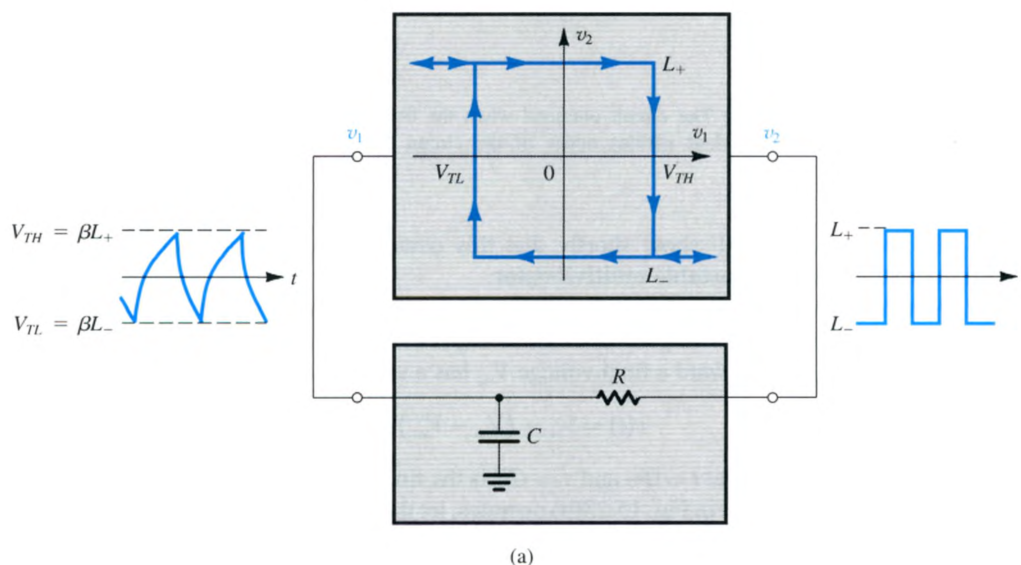


Figure 15.23 **(a)** Connecting a bistable multivibrator with inverting transfer characteristics in a feedback loop with an RC circuit results in a square-wave generator.

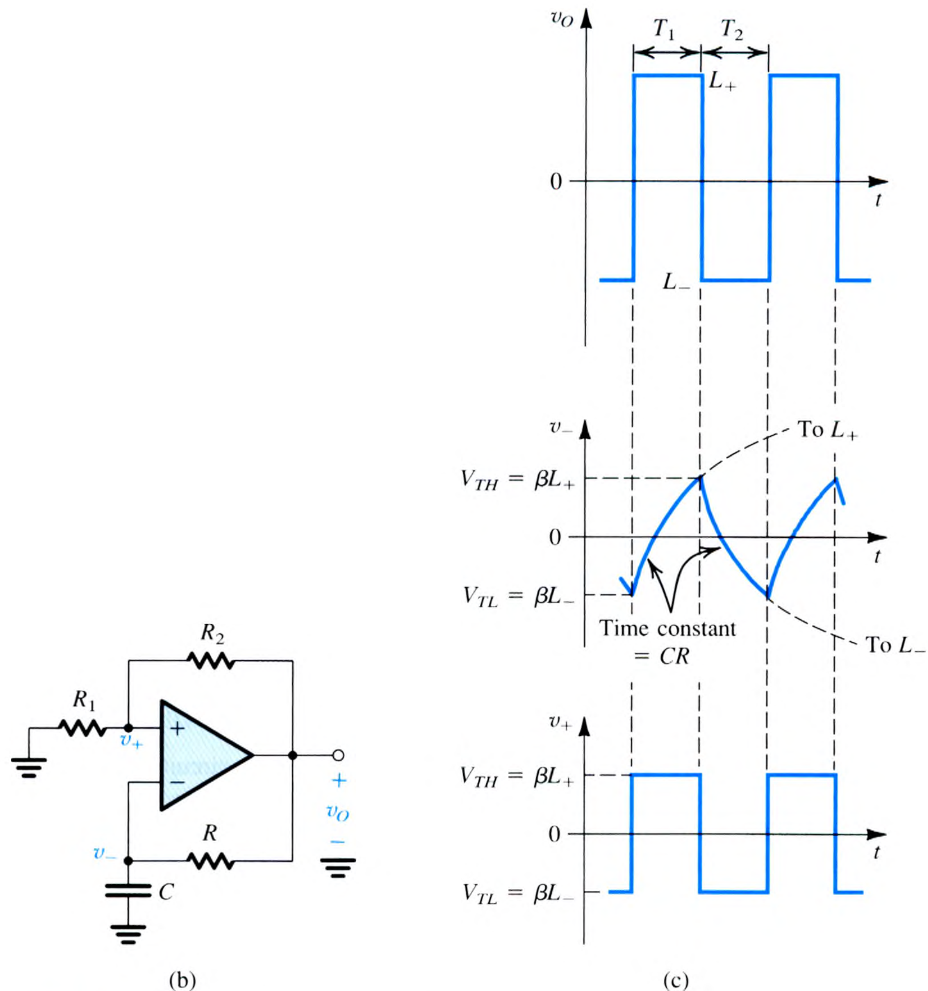


Figure 15.23 *continued* **(b)** The circuit obtained when the bistable is implemented with the circuit of Fig. 15.21(a). **(c)** Waveforms at various nodes of the circuit in **(b)**. This circuit is called an astable multivibrator.

of Fig. 15.23(b). We will show shortly that this circuit has no stable states and thus is appropriately named an **astable multivibrator**.

At this point we wish to remind you of an important relationship, which we will use several times in the following discussion: A capacitor C that is charging or discharging through a resistance R toward a final voltage V_∞ has a voltage $v(t)$,

$$v(t) = V_\infty - \left(V_\infty - V_{0+}\right)e^{-t/\tau}$$

where V_{0+} is the voltage at $t = 0+$ and $\tau = CR$ is the time constant.

To see how the circuit in Fig. 15.23(b) operates, let the output of the bistable multivibrator be at one of its two possible levels, say L_+. Capacitor C will charge toward this level through resistor R. Thus the voltage across C, which is applied to the negative input terminal of the op amp and thus is denoted v_-, will rise exponentially toward L_+ with a time constant $\tau = CR$. Meanwhile, the voltage at the positive input terminal of the op amp is $v_+ = \beta L_+$

(where $\beta = R_1/(R_1 + R_2)$). This situation will continue until the capacitor voltage reaches the positive threshold $V_{TH} = \beta L_+$, at which point the bistable multivibrator will switch to the other stable state, in which $v_O = L_-$ and $v_+ = \beta L_-$. The capacitor will then start discharging, and its voltage, v_-, will decrease exponentially toward L_-. This new state will prevail until v_- reaches the negative threshold $V_{TL} = \beta L_-$, at which time the bistable multivibrator switches to the positive-output state, the capacitor begins to charge, and the cycle repeats itself.

From the preceding description we see that the astable circuit oscillates and produces a square waveform at the output of the op amp. This waveform, and the waveforms at the two input terminals of the op amp, are displayed in Fig. 15.23(c). The period T of the square wave can be found as follows: During the charging interval T_1 the voltage v_- across the capacitor at any time t, with $t = 0$ at the beginning of T_1, is given by

$$v_- = L_+ - (L_+ - \beta L_-)e^{-t/\tau}$$

where $\tau = CR$. Substituting $v_- = \beta L_+$ at $t = T_1$ gives

$$T_1 = \tau \ln \frac{1 - \beta(L_-/L_+)}{1 - \beta} \tag{15.40}$$

Similarly, during the discharge interval T_2 the voltage v_- at any time t, with $t = 0$ at the beginning of T_2, is given by

$$v_- = L_- - (L_- - \beta L_+)e^{-t/\tau}$$

Substituting $v_- = \beta L_-$ at $t = T_2$ gives

$$T_2 = \tau \ln \frac{1 - \beta(L_+/L_-)}{1 - \beta} \tag{15.41}$$

Equations (15.40) and (15.41) can be combined to obtain the period $T = T_1 + T_2$. Normally, $L_+ = -L_-$, resulting in a symmetrical square wave of period T given by

$$T = 2\tau \ln \frac{1 + \beta}{1 - \beta} \tag{15.42}$$

This square-wave generator can be made to have variable frequency by switching different capacitors C (usually in decades) and continuously adjusting R (to obtain continuous frequency control within each decade of frequency). Also, the waveform across C can be made almost triangular by using a small value for the parameter β. However, triangular waveforms of superior linearity can be easily generated using the scheme discussed next.

Before leaving this section, note that although the astable circuit has no stable states, it has two *quasi-stable* states and remains in each for a time interval determined by the time constant of the RC network and the thresholds of the bistable multivibrator.

EXERCISES

15.17 For the circuit in Fig. 15.23(b), let the op-amp saturation voltages be ±10 V, $R_1 = 100$ kΩ, $R_2 = R = 1$ MΩ, and $C = 0.01$ μF. Find the frequency of oscillation.
Ans. 274 Hz

15.18 Consider a modification of the circuit of Fig. 15.23(b) in which R_1 is replaced by a pair of diodes connected in parallel in opposite directions. For $L_+ = -L_- = 12$ V, $R_2 = R = 10$ kΩ, $C = 0.1$ μF, and the diode voltage as a constant denoted V_D, find an expression for frequency as a function of V_D. If $V_D = 0.70$ V at 25°C with a TC of –2 mV/°C, find the frequency at 0°C, 25°C, 50°C, and 100°C. Note that the output of this circuit can be sent to a remotely connected frequency meter to provide a digital readout of temperature.
Ans. $f = 500/\ln[(12 + V_D)/(12 - V_D)]$ Hz; 3995 Hz, 4281 Hz, 4611 Hz, 5451 Hz

15.4.7 Generating Triangular Waveforms

The exponential waveforms generated in the astable circuit of Fig. 15.23 can be changed to triangular by replacing the low-pass RC circuit with an integrator. (The integrator is, after all, a low-pass circuit with a corner frequency at dc.) The integrator causes linear charging and discharging of the capacitor, thus providing a triangular waveform. The resulting circuit is shown in Fig. 15.24(a). Observe that because the integrator is inverting, it is necessary to invert the characteristics of the bistable circuit. Thus the bistable circuit required here is of the noninverting type and can be implemented using the circuit of Fig. 15.22(a).

We now proceed to show how the feedback loop of Fig. 15.24(a) oscillates and generates a triangular waveform v_1 at the output of the integrator and a square waveform v_2 at the output of the bistable circuit: Let the output of the bistable circuit be at L_+. A current equal to L_+/R will flow into the resistor R and through capacitor C, causing the output of the integrator to *linearly* decrease with a slope of $-L_+/CR$, as shown in Fig. 15.24(c). This will continue until the integrator output reaches the lower threshold V_{TL} of the bistable circuit, at which point the bistable circuit will switch states, its output becoming negative and equal to L_-. At this moment the current through R and C will reverse direction, and its value will become equal to $|L_-|/R$. It follows that the integrator output will start to increase linearly with a positive slope equal to $|L_-|/CR$. This will continue until the integrator output voltage reaches the positive threshold of the bistable circuit, V_{TH}. At this point the bistable circuit switches, its output becomes positive (L_+), the current into the integrator reverses direction, and the output of the integrator starts to decrease linearly, beginning a new cycle.

From the discussion above, it is relatively easy to derive an expression for the period T of the square and triangular waveforms. During the interval T_1 we have, from Fig. 15.24(c),

$$\frac{V_{TH} - V_{TL}}{T_1} = \frac{L_+}{CR}$$

from which we obtain

$$T_1 = CR\frac{V_{TH} - V_{TL}}{L_+} \tag{15.43}$$

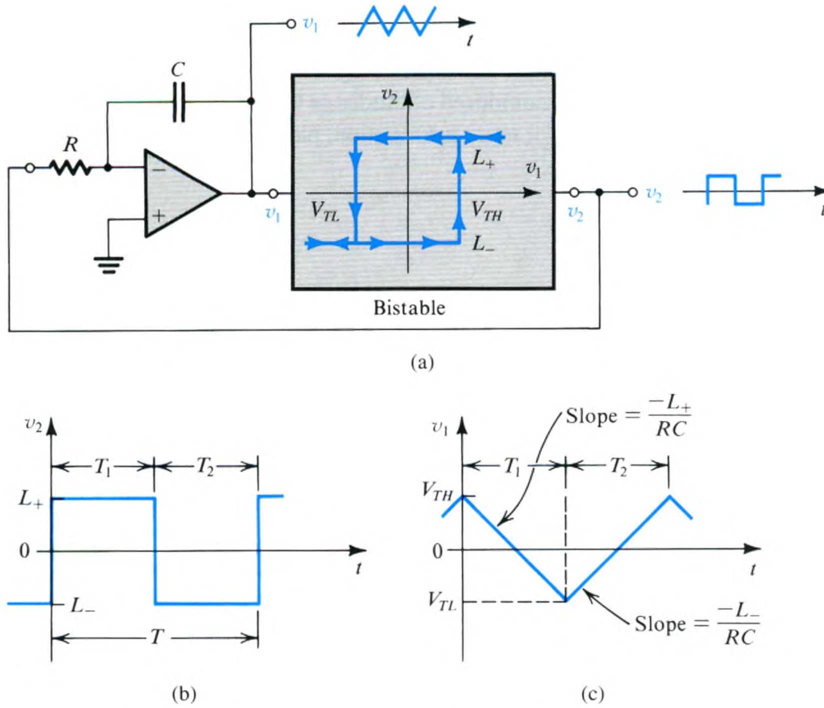


Figure 15.24 A general scheme for generating triangular and square waveforms. The circuit in Fig. 15.22(a) can be used to implement the non-inverting bistable.

Similarly, during T_2 we have

$$\frac{V_{TH} - V_{TL}}{T_2} = \frac{-L_-}{CR}$$

from which we obtain

$$T_2 = CR\frac{V_{TH} - V_{TL}}{-L_-} \tag{15.44}$$

Thus to obtain symmetrical square waves [Fig. 15.24(b)] and symmetrical triangular waves [Fig. 15.24(c)], we design the bistable circuit to have $L_+ = -L_-$.

EXERCISE

D15.19 Consider the circuit of Fig. 15.24(a) with the bistable circuit realized with the circuit in Fig. 15.22(a). If the op amps have saturation voltages of ± 10 V, and if a capacitor $C = 0.01\ \mu\text{F}$ and a resistor $R_1 = 10\ \text{k}\Omega$ are used, find the values of R and R_2 (note that R_1 and R_2 are associated with the bistable circuit of Fig. 15.22a) such that the frequency of oscillation is 1 kHz and the triangular waveform has a 10-V peak-to-peak amplitude.

Ans. 50 kΩ; 20 kΩ

15.4.8 Generation of Sine Waves

In a function generator, sine waveforms can be obtained by feeding triangular waves to a "sine-wave shaper." The latter is a circuit composed of diodes or transistors and resistors. We will not discuss sine-wave shapers here; for more information, please refer to the companion website: www.oup.com/he/sedra-smith8e.

Summary

- There are two distinctly different types of signal generator: the linear oscillator, which utilizes some form of resonance, and the nonlinear oscillator or function generator, which employs a switching mechanism implemented with a multivibrator circuit.
- A linear oscillator can be realized by placing a frequency-selective network (that is, a filter) in the feedback path of an amplifier (an op amp or a transistor). The circuit will oscillate at the frequency at which the total phase shift around the loop is zero or 360°, provided the magnitude of loop gain at this frequency is equal to, or greater than, unity.
- If in an oscillator the magnitude of loop gain is greater than unity, the amplitude will increase until a nonlinear amplitude-control mechanism is activated.
- The Wien-bridge oscillator, the phase-shift oscillator, the quadrature oscillator, and the active-filter-tuned oscillator are popular configurations for generating sine waves with frequencies up to about 1 MHz. These circuits employ RC networks together with op amps or transistors. For higher frequencies, LC-tuned or crystal-tuned oscillators are utilized. Popular configurations include the Colpitts circuit for discrete-circuit implementation and the cross-coupled circuit for IC implementation at frequencies as high as hundreds of gigahertz.
- Crystal oscillators provide the highest possible frequency accuracy and stability.
- The bistable multivibrator has two stable states and can remain in either state indefinitely. It changes state when triggered. A comparator with hysteresis is bistable.
- An astable multivibrator has no stable state. It oscillates between two quasi-stable states, remaining in each for a predetermined interval. It thus generates a periodic waveform at the output.
- A feedback loop consisting of an integrator and a bistable multivibrator can be used to generate triangular and square waveforms.

PROBLEMS

Section 15.1: Basic Principles of Sinusoidal Oscillators

15.1 A sinusoidal oscillator consists of an amplifier having a positive, frequency-independent gain A connected in a feedback loop with a second-order bandpass filter having the transfer function

$$T(s) = \frac{sK(\omega_0/Q)}{s^2 + s(\omega_0/Q) + \omega_0^2}$$

where K is positive. Find the frequency of oscillation and the minimum value that K must have for sustained oscillations.

15.2 For the oscillator described in Problem 15.1, find the characteristic equation and use it to show that, independent of the values of A and K, the poles of the circuit lie at a radial distance of ω_0. Find the values of AK that results in the poles appearing (a) on the $j\omega$ axis, and (b) in the right half of the s plane, at a horizontal distance from the $j\omega$ axis of $\omega_0/(2Q)$.

D 15.3 The characteristic equation of the oscillator circuit of Fig. 15.2(a) is derived in Example 15.2 and given by Eq. (15.6):

$$s^2 - s\frac{1}{CR}\left(\frac{r_2}{r_1}\right) + \frac{1}{LC} = 0$$

problems with blue numbers are considered essential; * = difficult problem; ** = more difficult; *** = very challenging
SIM = simulation; D = design problem; ● = see related video example

Note that while $r_2 = 0$ results in sustained oscillation, making $r_2 > 0$ places the poles in the right half of the s-plane and ensures that oscillation will start and grow until a nonlinear amplitude-control mechanism kicks in. Using an inductance of 0.1 mH, find

(a) the value of C to obtain sinusoidal oscillations with 10^7 rad/s frequency;
(b) the value of R so that the LCR circuit has a Q factor of 50 (note: a high Q-factor, $Q = \omega_0 CR$, increases the purity of the generated sine waves); and
(c) the value of (r_2/r_1) to place the poles at a horizontal distance of 100 rad/s from the $j\omega$-axis.

15.4 In a particular oscillator characterized by the structure of Fig. 15.1, the frequency-selective network exhibits a loss of 12 dB and a phase shift of 180° at ω_0. Give the phase shift and the minimum gain that the amplifier must have for oscillation to begin.

15.5 An oscillator is formed by loading a transconductance amplifier having a positive gain with a parallel RLC circuit and connecting the output directly to the input (thus applying positive feedback with a factor $\beta = 1$). Let the transconductance amplifier have an input resistance of 5 kΩ and an output resistance of 5 kΩ. The LC resonator has $L = 1\ \mu$H, $C = 100$ pF, and $Q = 50$. For what value of transconductance G_m will the circuit oscillate? At what frequency?

15.6 Figure P15.6 shows a sinusoidal oscillator using a simple RC bandpass circuit in the feedback loop of a positive-gain amplifier. Find the loop gain as a function of s and, by substituting $s = j\omega$, as a function of frequency ω. What is the frequency of oscillation ω_0? What is the minimum value (r_2/r_1) must have for sustained oscillation?

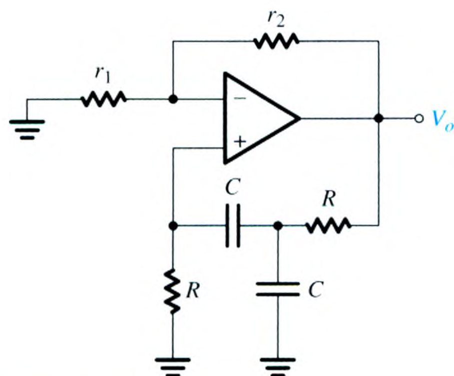


Figure P15.6

15.7 For the oscillator circuit in Fig. P15.6, use the alternative analysis method to determine the oscillation frequency ω_0 and the minimum required value of r_2/r_1. Specifically, assume that the circuit is oscillating and write equations that describe its operation. Reduce the equations to one in terms of V_o and use it to determine the conditions of oscillation.

15.8 Find the characteristic equation of the circuit in Fig. P15.6 and use it to determine the value (r_2/r_1) must have for the poles to lie on the $j\omega$-axis. This is the value required for sustained oscillations. What is the frequency of oscillation?

15.9 An oscillator is designed by connecting in a loop three identical common-source amplifier stages of the type shown in Fig. P15.9. Note that the bias circuits are not shown, and assume that R and C include the transistor output resistance and capacitance, respectively. For the circuit to oscillate at a frequency ω_0, what must the phase angle provided by each amplifier stage be? Give an expression for ω_0. For sustained oscillations, what is the minimum g_m required of each transistor?

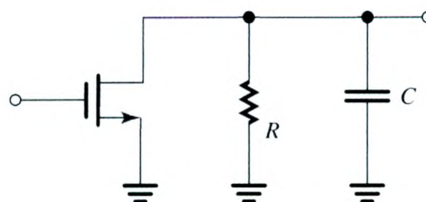


Figure P15.9

15.10 Assume that the oscillator of Fig. 15.2(a) is designed for a loop-gain magnitude greater than one so that the amplitude of oscillations grows to the point that the op-amp output saturates. Assume that the saturation levels of the op amp are ±2 V and that the signal at the output of the op amp is nearly a square wave. If the bandpass LCR circuit is of sufficiently high selectivity, what is the amplitude of the sine wave across the LC circuit? (*Hint:* The fundamental-frequency component in the Fourier series expansion of a square wave of amplitude V has an amplitude $4V/\pi$.)

15.11 For the oscillator circuit in Fig. 15.5, assume that v_L is an ideal square wave of frequency ω_0 and amplitude V volts. Recalling that the Fourier series for a square wave is

$$v_L = \frac{4V}{\pi}\left[\sin(\omega_0 t) + \frac{1}{3}\sin(3\,\omega_0 t) + \frac{1}{5}\sin(5\,\omega_0 t) + \cdots\right]$$

find the third harmonic component of the output v_O as a percentage of the fundamental component. Repeat for the

fifth harmonic. Assume the filter has $Q = 100$. (*Hint:* The filter transfer function is

$$T(s) = \frac{s\left(\frac{\omega_0}{Q}\right)}{s^2 + s\left(\frac{\omega_0}{Q}\right) + \omega_0^2}$$

where $\omega_0 = 1/\sqrt{LC}$ and $Q = \omega_0 CR$.)

D 15.12 Using an inductance L = 0.1 mH, design the oscillator circuit in Fig. 15.5 to obtain a sinusoid at the amplifier output of 3-V amplitude and 100-kHz frequency. Assume that v_L is a perfect square wave with ±0.7-V levels. Use r_1 = 10 kΩ and design the filter to have a Q factor of 50. Specify the values of C, R, and r_2. What is the value of r_3 that limits the maximum diode current to 1 mA? (*Hint:* $Q = \omega_0 CR$.)

Section 15.2: Op Amp–RC Oscillator Circuits

15.13 For the Wien-bridge oscillator circuit in Fig. 15.6, show that the transfer function of the feedback network $[V_a(s)/V_o(s)]$ is that of a bandpass filter. Find ω_0 and Q of the poles, and find the center-frequency gain. Hence, find the oscillation frequency and the value of (R_2/R_1) for sustained oscillations.

15.14 For the Wien-bridge oscillator of Fig. 15.6, let the closed-loop amplifier (formed by the op amp and the resistors R_1 and R_2) exhibit a phase shift of −5° in the neighborhood of $\omega = 1/CR$. Find the frequency at which oscillations can occur in this case in terms of CR. (*Hint:* Use a suitably modified version of Eq. 15.8.)

15.15 For the Wien-bridge oscillator of Fig. 15.6, use the expression for loop gain in Eq. (15.7) to find the characteristic equation of the closed-loop system. Show that to locate the poles on the $j\omega$-axis, R_2/R_1 must be selected equal to 2. What is the frequency of oscillation?

D 15.16 Reconsider Exercise 15.5 with R_3 and R_6 increased to reduce the output voltage. What values are required for a peak-to-peak output of 10 V? What results if R_3 and R_6 are open-circuited?

D 15.17 Design the Wien-bridge oscillator using a limiter similar to the one used in Fig. 15.7 for amplitude control. Find suitable component values to obtain a sinusoid v_O with 10-kHz frequency and 3-V peak amplitude. Assume V_{DD} = V_{SS} = 5 V and V_D = 0.7 V. Use as many 10-kΩ resistors as possible.

15.18 For the circuit in Fig. P15.18, find $L(s)$, $L(j\omega)$, the frequency for zero loop phase, and R_2/R_1 for oscillation. Assume the op amp to be ideal.

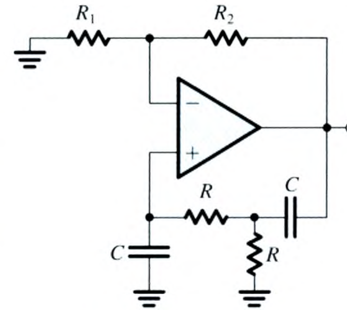


Figure P15.18

***15.19** Consider the circuit of Fig. 15.8 with the 50-kΩ potentiometer replaced with two fixed resistors: 10 kΩ between the op amp's negative input and ground, and 15 kΩ. Modeling each diode as a 0.65-V battery in series with a 100-Ω resistance, find the peak-to-peak amplitude of the output sinusoid.

D **15.20 Design the Wien-bridge oscillator for operation at 10 kHz using R = 10 kΩ. If at 10 kHz the op amp provides an excess phase shift (lag) of 5.7°, what will be the frequency of oscillation? (Assume that the phase shift introduced by the op amp remains constant for frequencies around 10 kHz.) To restore operation to 10 kHz, what change must be made in the shunt resistor of the Wien bridge? Also, to what value must R_2/R_1 be changed?

***15.21** For the circuit of Fig. 15.10, connect an additional resistor (R = 10 kΩ) in series with the rightmost capacitor C. For this modification (and ignoring the amplitude stabilization circuitry), find the loop gain $A\beta$ by breaking the circuit at node X. Find R_f for oscillation to begin, and find f_0. Do the analysis symbolically and substitute the given values only at the end.

D 15.22 For the circuit in Fig. P15.22, break the loop at node X and find the loop gain (working backward for simplicity to find V_x in terms of V_o). Find expressions for

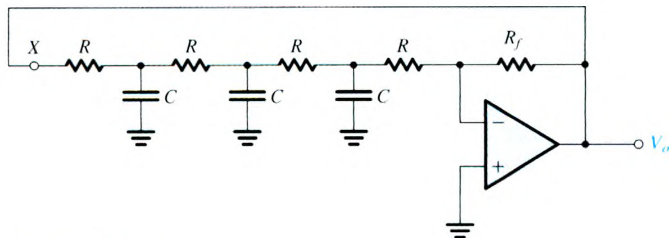


Figure P15.22

ω_0 and R_f for sustained oscillations. For $R = 10$ kΩ, find C and R_f to obtain sinusoidal oscillations at 20 kHz.

15.23 For the circuit in Fig. 15.10, let $V^+ = -V^- = V$, $R_4 = R_1$, and $R_3 = R_2$. Assume that a conducting diode has a voltage drop V_D and that at the onset of conduction the diode current can be neglected relative to the current through R_1, R_2, R_3, and R_4. Show that the peak amplitude of v_O is given by

$$\hat{v}_O = V_D + \left(\frac{R_3}{R_4}\right)(V + V_D)$$

For $V = 5$ V, $V_D = 0.7$ V, find R_3/R_4 to obtain $\hat{v}_O = 2$ V.

***15.24** Consider the quadrature-oscillator circuit of Fig. 15.11 without the limiter. Let the resistance R_f be equal to $2R/(1+\Delta)$, where $\Delta \ll 1$. Show that the poles of the characteristic equation are in the right-half s plane and given by $s \simeq (1/CR)\left[(\Delta/4) \pm j\right]$.

15.25 Derive the transfer function V_{o2}/V_{o1} of the noninverting integrator formed by op amp 2 and the associated components in the circuit of Fig. 15.11(a). Show that to obtain a perfect integrator, R_f must be chosen equal to $2R$.

D 15.26 Design the quadrature oscillator circuit in Fig. 15.11(a) to obtain sinusoids with 5-V amplitude and 10-kHz frequency. Use $R = 10$ kΩ, $V^+ = -V^- = 15$ V, $R_1 = R_4 = 10$ kΩ, $R_2 = R_3$, and the diode voltage drop $V_D = 0.7$ V. Specify all component values.

15.27 Derive the transfer function V_1/V_2 of the bandpass circuit that is part of the oscillator in Fig. 15.13. Show that its poles have $\omega_0 = 1/CR$ and a Q factor Q, and that the center-frequency gain is 2.

D 15.28 Using $C = 1.6$ nF, find the value of R such that the circuit of Fig. 15.13 produces 20-kHz sine waves. If the diode drop is 0.7 V, find the peak-to-peak amplitude of the output sine wave. How do you modify the circuit to double the output amplitude? (*Hint*: A square wave with peak-to-peak amplitude of V volts has a fundamental component with $4V/\pi$ volts peak-to-peak amplitude.)

***15.29** Assuming that the diode-clipped waveform in Exercise 15.9 is an ideal square wave and that the resonator Q is 30, provide an estimate of the distortion in the output sine wave by calculating the magnitude (relative to the fundamental) of

(a) the second harmonic
(b) the third harmonic
(c) the fifth harmonic
(d) the rms of harmonics to the tenth

Note that a square wave of amplitude V and frequency ω is represented by the Fourier series

$$\frac{4V}{\pi}\left(\sin\omega t + \frac{1}{3}\sin 3\omega t + \frac{1}{5}\sin 5\omega t + \frac{1}{7}\sin 7\omega t + \ldots\right)$$

Section 15.3: LC and Crystal Oscillators

15.30 For the Colpitts oscillator circuit in Fig. P15.30, derive an equation governing circuit operation and hence find the frequency of oscillation and the condition the gain $g_m R_L$ must satisfy for oscillations to start. Assume that R_L includes the MOSFET's r_o. Note that this circuit is based on the configuration in Fig. 15.14(a).

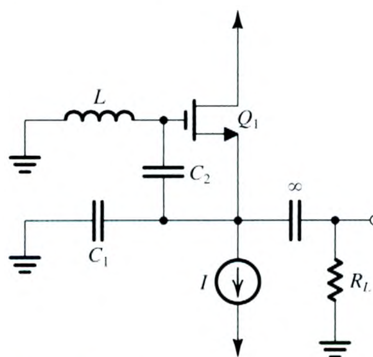


Figure P15.30

15.31 For the Colpitts oscillator circuit in Fig. P15.31, derive an equation governing circuit operation and hence find

the frequency of oscillation and the condition the gain $g_m R_L$ must satisfy to ensure that oscillations will start. Neglect r_o of the MOSFET. Note that this circuit is based on the configuration in Fig. 15.14(a).

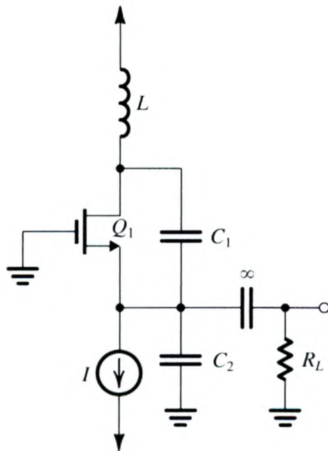


Figure P15.31

15.32 For the Colpitts oscillator circuit in Fig. P15.32, derive an equation governing circuit operation and hence find the frequency of oscillation and the condition the gain $g_m R_L$ must satisfy to ensure that oscillations will start. Assume that r_o of the MOSFET is included in R_L and neglect R_f (i.e., assume $R_f \gg \omega_0 L$). Observe that this circuit is based on the configuraion in Fig. 15.14(a).

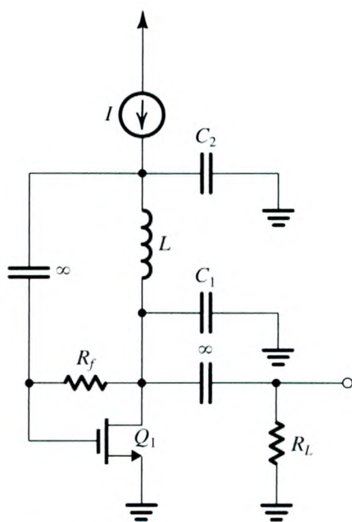


Figure P15.32

D 15.33 Design the Hartley oscillator in Fig. 15.14(b) to obtain sine waves of 10-MHz frequency. Assume the MOSFET is biased to operate at $g_m = 2$ mA/V and that $r_o = 20$ kΩ. Let the finite Q of the inductors be represented by a 20 kΩ resistance across L_1. Use $C = 100$ pF and neglect the capacitances of the MOSFET. Find L_1 and L_2.

***15.34** The LC oscillator in Fig. P15.34 is based on connecting a positive-gain amplifier (formed by Q_1, Q_2, and R_C) with a bandpass RLC circuit in a feedback loop.

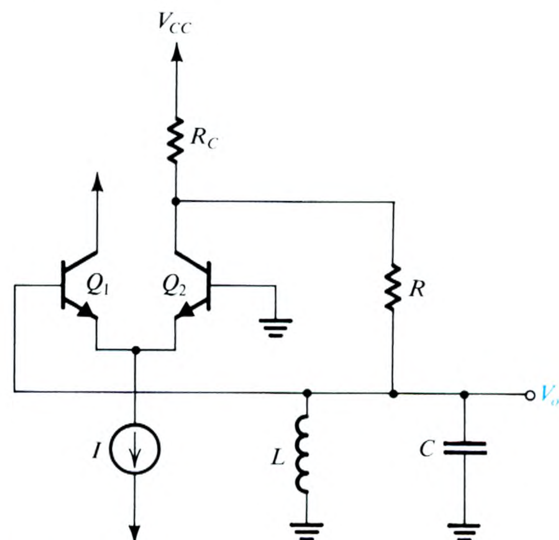


Figure P15.34

(a) Replace the BJTs with their small-signal models while neglecting r_π and r_o (to simplify matters).
(b) By inspection of the circuit found in (a), find the frequency of oscillation and the condition required for oscillations to start. Express the latter as the minimum required value of (IR_C).
(c) If IR_C is selected equal to 1 V, show that oscillations will start. If oscillations grow to the point that V_o is large enough to turn the BJTs on and off, show that the signal at the collector of Q_2 will be a square wave of 1 V peak-to-peak. Estimate the peak-to-peak amplitude of the output sine wave V_o.

problems with blue numbers are considered essential; * = difficult problem; ** = more difficult; *** = very challenging
SIM = simulation; D = design problem; ● = see related video example

(*Hint:* A square wave with amplitude V has a fundamental-frequency component with amplitude $4V/\pi$.)

D15.35 Design the cross-coupled LC oscillator of Fig. 15.16(a) to operate at $\omega_0 = 50$ Grad/s. The IC inductors available have $L = 2$ nH and $Q = 10$. If the transistor $r_o = 5$ kΩ, find the required value of C and the minimum required value of g_m at which Q_1 and Q_2 are to be operated.

15.36 Consider the Pierce crystal oscillator of Fig. 15.18 with the crystal as specified in Exercise 15.13. Let C_1 be variable in the range 1 pF to 10 pF, and let C_2 be fixed at 10 pF. Find the range over which the oscillation frequency can be tuned. (*Hint:* Use the result in the statement leading to the expression in Eq. 15.36.)

Section 15.4: Nonlinear Oscillators or Function Generators

D 15.37 Design the bistable circuit in Fig. 15.21(a) to obtain a hysteresis of 2.5-V width. The op amp saturates at ±5 V. Select $R_1 = 10$ kΩ and determine R_2.

15.38 Consider the bistable circuit of Fig. 15.21(a) with the op amp's positive input terminal connected to a positive-voltage source V through a resistor R_3.

(a) Derive expressions for the threshold voltages V_{TL} and V_{TH} in terms of the op amp's saturation levels L_+ and L_-, R_1, R_2, R_3, and V.
(b) Let $L_+ = -L_- = 5$ V, $V = 5$ V, and $R_1 = 10$ kΩ. Find the values of R_2 and R_3 that result in $V_{TL} = +1.9$ V and $V_{TH} = +2.1$ V.

D 15.39 For the circuit in Fig. 15.22(a), let $L_+ = -L_- = 5$ V and $R_1 = 10$ kΩ. Find R_2 that results in threshold voltages of +1 V and −1 V.

15.40 Consider the bistable circuit of Fig. 15.22(a) with the op amp's negative-input terminal disconnected from ground and connected to a reference voltage V_R.

(a) Derive expressions for the threshold voltages V_{TL} and V_{TH} in terms of the op amp's saturation levels L_+ and L_-, R_1, R_2, and V_R.
(b) Let $L_+ = -L_- = V$ and $R_1 = 10$ kΩ. Find R_2 and V_R that result in threshold voltages of 0 and $V/10$.

15.41 For the circuit in Fig. P15.41, sketch and label the transfer characteristic v_O–v_I. The diodes are assumed to have a constant 0.7-V drop when conducting, and the op amp saturates at ±5 V. What is the maximum diode current?

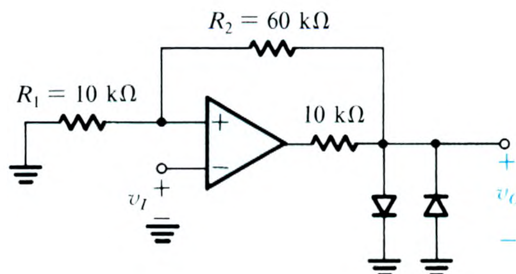


Figure P15.41

15.42 Consider the circuit of Fig. P15.41 with R_1 eliminated and R_2 short-circuited. Sketch and label the transfer characteristic v_O–v_I. Assume that the diodes have a constant 0.7-V drop when conducting and that the op amp saturates at ±5 V.

***15.43** Consider a bistable circuit having a noninverting transfer characteristic with $L_+ = -L_- = 5$ V, $V_{TL} = -1$ V, and $V_{TH} = +1$ V.

(a) For a 0.5-V-amplitude sine-wave input having zero average, what is the output?
(b) Describe the output if a sinusoid of frequency f and amplitude of 1.1 V is applied at the input. By how much can the average of this sinusoidal input shift before the output becomes a constant value?

15.44 Find the frequency of oscillation of the circuit in Fig. 15.23(b) for the case $R_1 = 10$ kΩ, $R_2 = 15$ kΩ, $C = 1$ nF, and $R = 62$ kΩ.

D 15.45 Design the circuit in Fig. 15.23(b) to obtain a square waveform of 10-kHz frequency. Use $C = 1$ nF and $R_1 = R_2$. Specify the values of the three resistors.

D 15.46 Design the circuit in Fig. 15.24(a) to obtain symmetrical triangular and square waveforms of 10-kHz frequency. Use the circuit in Fig. 15.22(a) to implement the bistable. Assume that the op amps saturate at ±5 V. The triangular waveform is required to have 2-V peak-to-peak amplitude. Use $C = 1$ nF and specify the values of the resistors.

PART III

Digital Integrated Circuits

There are two indisputable facts about digital systems. They have dramatically changed our lives, and the digital revolution is driven by microelectronics.

Evidence of the pervasiveness and influence of digital systems can be found by thinking of what we do in our daily lives. Digital circuits exist in almost every electrical appliance we use in our homes; in the vehicles and transportation systems we use to travel; in the smartphones and tablets we use to communicate; in the medical equipment needed to care for our health; in the computers we use to do our work; in our entertainment systems; and in the servers running cutting-edge machine-learning algorithms. Indeed, it is very difficult to conceive of modern life without digital systems, none of which would have been possible without microelectronics.

Although the idea of a digital computing machine was conceived as early as the 1830s, early implementations were very cumbersome and used expensive mechanical devices. The first serious digital computers appeared in the 1930s and 1940s. These early computers used thousands of vacuum tubes and were housed literally in many rooms. Their fundamental limitation was low reliability: vacuum tubes had a finite life and needed large amounts of power. Had it not been for the invention of the transistor in 1947, ushering in the era of solid-state electronics, digital computers would have remained specialized machines used primarily in military and scientific applications.

By the mid-1950s, the first digital logic gates made of discrete bipolar transistors were commercially available. The invention of the integrated circuit in the late 1950s was also key, leading to the first digital IC in the early 1960s. Early digital ICs were made of bipolar transistors, with the most successful logic-circuit family of this type being transistor–transistor logic (or TTL), which dominated digital-circuit design until the early 1980s.

Bipolar was replaced by NMOS, and NMOS by CMOS, again predominantly because of power dissipation and the need to pack more and more transistors on each IC chip. Bearing out Moore's law, which predicted in 1968 that IC chips would double the number of their transistors every two to three years (see Section 17.4), digital ICs have grown from a few transistors to tens of billions of devices and to memory chips with 64-Gbit capacity.

Part III aims to provide a concise but nonetheless comprehensive and sufficiently detailed exposure to digital IC design. Our treatment is almost self-contained, requiring for the most part only a thorough understanding of the MOSFET material presented in Chapter 5. Thus Part III can be studied right after Chapter 5. The only exception to this is that Chapters 17 and 18 require knowledge of the MOSFET internal capacitances (Section 10.1).

Chapter 16 is the cornerstone of Part III. It provides a study of the bread-and-butter topic of digital IC design: the CMOS inverter and logic gates. Today, CMOS represents 98% of newly designed digital systems. The material in Chapter 16 is the minimum needed to learn something meaningful about digital circuits and their static behaviour; it is a must study!

In Chapter 17, we examine the dynamic behavior of CMOS logic circuits. This serves as an introduction to several interrelated techniques to optimize speed, power dissipation, and silicon area. Chapter 17 concludes with a discussion of transistor scaling and future technologies, with emphasis on performance issues and economic aspects.

Digital circuits can be broadly divided into logic and memory circuits. The latter form the subject of Chapter 18, which begins with a discussion of transmission gates, and then covers latches, flip-flops, as well as static and dynamic memory cells and their operation.

CHAPTER 16

CMOS Digital Logic Circuits

IN THIS CHAPTER YOU WILL LEARN

- How CMOS logic circuits can be implemented using arrangements of MOS transistors operating as voltage-controlled switches.
- How to synthesize CMOS logic circuits that realize standard (e.g., inverter, NAND, and NOR gates) and complex Boolean functions.
- How the inverter can be implemented by using one of three possible arrangements of voltage-controlled switches (transistors).
- The structure, circuit operation, and design of the CMOS inverter, as well as analysis of its static behavior.

Introduction

This chapter provides a foundation for the study of CMOS logic circuits, the bread and butter of digital IC design. We begin (Section 16.1) by learning how to synthesize CMOS circuits that implement various logic functions. This discussion will be at a high level without getting into the details of circuit operation and performance. To delve into these issues, we consider in Section 16.2 the most fundamental element of digital circuits: the logic inverter. We study its characteristics, performance metrics, and methods of implementation. Out of this general study, CMOS emerges as the most ideal inverter implementation. Consequently, a thorough study of the CMOS inverter is undertaken in Section 16.3. Besides presenting the most important digital IC technology (CMOS), this chapter lays the foundation for the more advanced topics studied in the two subsequent chapters.

16.1 CMOS Logic-Gate Circuits

In this section we consider the synthesis of CMOS circuits that realize combinational-logic functions. In combinational circuits, the output at any time is a function only of the values of input signals at that time. Thus, these circuits do not have memory and do not employ feedback. Combinational circuits are abundant in every digital system.

16.1.1 Switch-Level Transistor Model

CMOS digital circuits utilize NMOS and PMOS transistors operating as switches. From Chapter 5, we know that a MOS transistor can operate as an on/off switch by using the gate voltage to operate the transistor in the triode region ("on" position) or in the cutoff region ("off" position).

Specifically, an NMOS transistor behaves as a closed switch, exhibiting a very small resistance (R_{on} or r_{DS}) between its drain and source terminals when its gate voltage is "high," usually at the power-supply level V_{DD}, which represents a logic 1. Conversely, when the gate voltage is "low" (i.e., at or close to ground voltage), which represents a logic 0, the transistor is cut off, thus conducting zero current and acting as an open switch. This is illustrated in Fig. 16.1(a).

The PMOS transistor operates in a complementary fashion: To turn the transistor on, its gate voltage is made low (0 V or logic 0). Raising the gate voltage to V_{DD} (logic 1) turns the PMOS transistor off. This is illustrated in Fig. 16.1(b).

We observe that the gate terminal of the MOSFET is used as the controlling node, and thus it is usually one of the input terminals of the logic gate.

16.1.2 The CMOS Inverter

Armed with this knowledge of the switching behavior of MOSFETs, let's consider making an inverter. As its name implies, the logic inverter inverts the logic value of its input signal. So, for a logic-0 input, the output will be a logic 1, and vice versa. Thus the logic function of the inverter can be represented by the Boolean expression

$$Y = \overline{X}$$

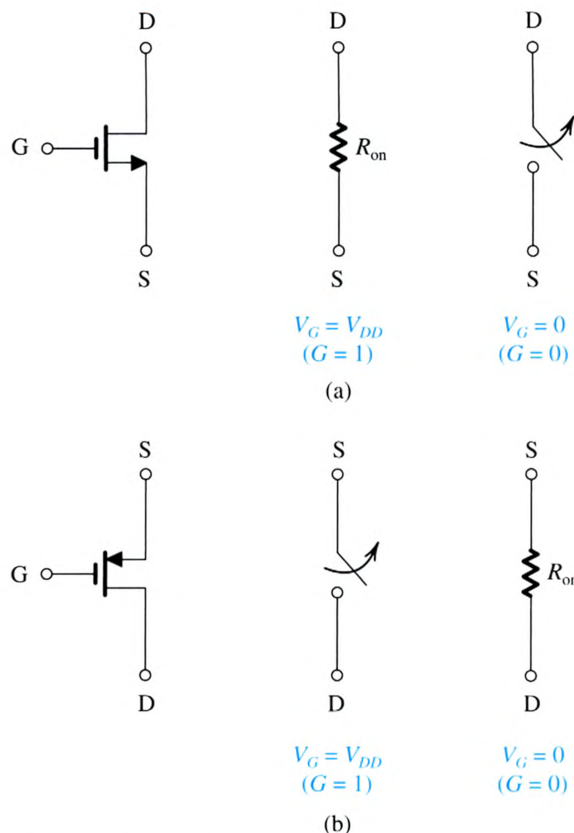


Figure 16.1 Operation of the (**a**) NMOS and (**b**) PMOS transistor as an on/off switch. The gate voltage controls the operation of the transistor switch, with the voltage V_{DD} representing a logic 1 and 0 V representing a logic 0. Note that the connections of the drain and source terminals are not shown.

An inverter operated from a power supply V_{DD} is shown in block form in Fig. 16.2(a). Its CMOS circuit implementation is shown in Fig. 16.2(b). It consists of an NMOS transistor Q_N and a PMOS transistor Q_P, with the gate terminals connected together to constitute the inverter input terminal, to which a logic input X is applied. Also, both drain terminals are connected together to constitute the inverter output terminal on which the output logic variable Y appears.

When $X = 1$—that is, $V_X = V_{DD}$ [Fig. 16.2(c)]—the PMOS transistor will be off but the NMOS transistor will be on and will be connecting the inverter output terminal to ground through the small on-resistance R_{on}. Thus, the output voltage will be zero and $Y = 0$. When $X = 0$, that is, $V_X = 0$ [see Fig. 16.2(d)], the NMOS transistor will be off but the PMOS transistor will be on and will be connecting the output terminal to V_{DD} through the small resistance R_{on}. Thus the output voltage will be equal to V_{DD} and Y will be 1.

16.1.3 General Structure of CMOS Logic

A CMOS logic circuit is in effect an extension, or a generalization, of the CMOS inverter: The inverter consists of an NMOS **pull-down transistor** and a PMOS **pull-up transistor**, operated by the input voltage in a complementary fashion. The CMOS logic gate consists

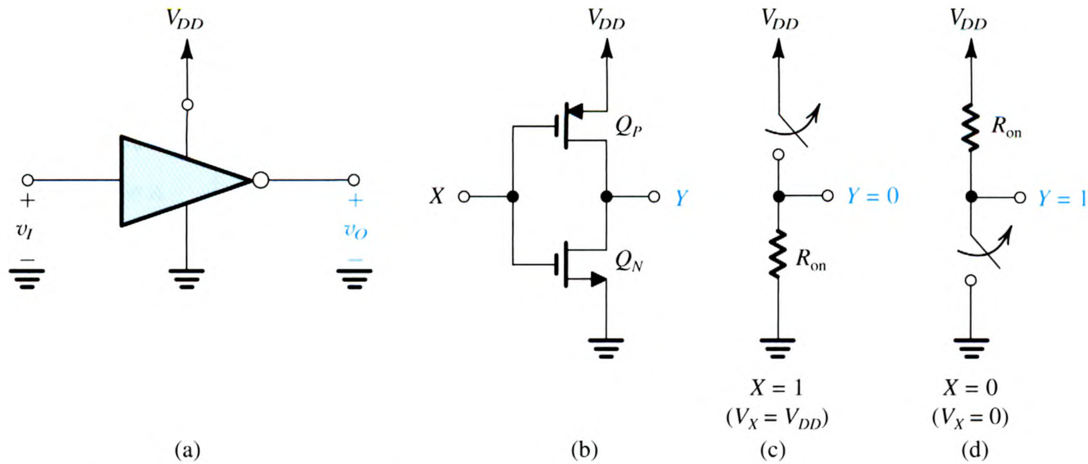


Figure 16.2 (**a**) Block representation of the logic inverter; (**b**) its CMOS realization; (**c**) operation when the input is a logic 1; (**d**) operation when the input is a logic 0.

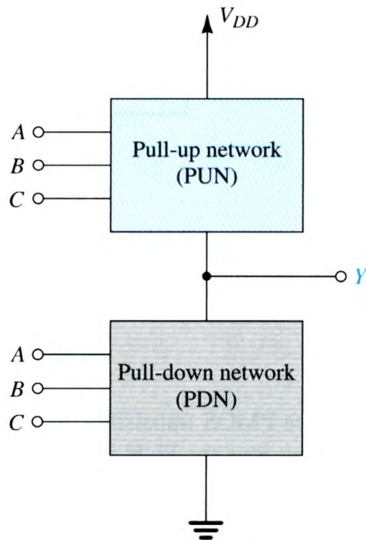


Figure 16.3 Representation of a three-input CMOS logic gate. The PUN comprises PMOS transistors, and the PDN comprises NMOS transistors.

of two networks: the **pull-down network (PDN)** constructed of NMOS transistors, and the **pull-up network (PUN)** constructed of PMOS transistors (see Fig. 16.3). The two networks are operated by the input variables, in a complementary fashion. Thus, for the three-input gate represented in Fig. 16.3, the PDN will conduct for all input combinations that require a low output ($Y = 0$) and will then pull the output node down to ground, causing a zero voltage to appear at the output, $v_Y = 0$. Simultaneously, the PUN will be off, and no direct dc path will exist between V_{DD} and ground. On the other hand, all input combinations that call for a high output ($Y = 1$) will cause the PUN to conduct, and the PUN will then pull the output node up to V_{DD}, establishing an output voltage $v_Y = V_{DD}$. Simultaneously, the PDN will be cut off, and again, no dc current path between V_{DD} and ground will exist in the circuit.

Now, since the PDN comprises NMOS transistors, and since an NMOS transistor conducts when the signal at its gate is high, the PDN is activated (i.e., conducts) when

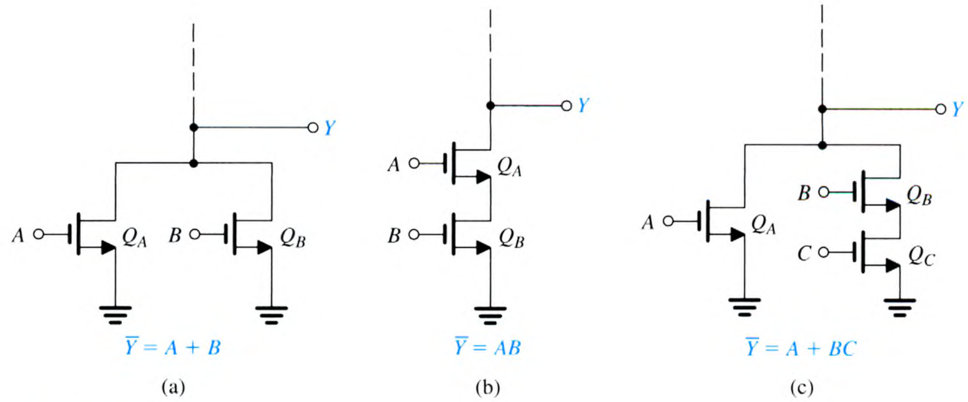


Figure 16.4 Examples of pull-down networks.

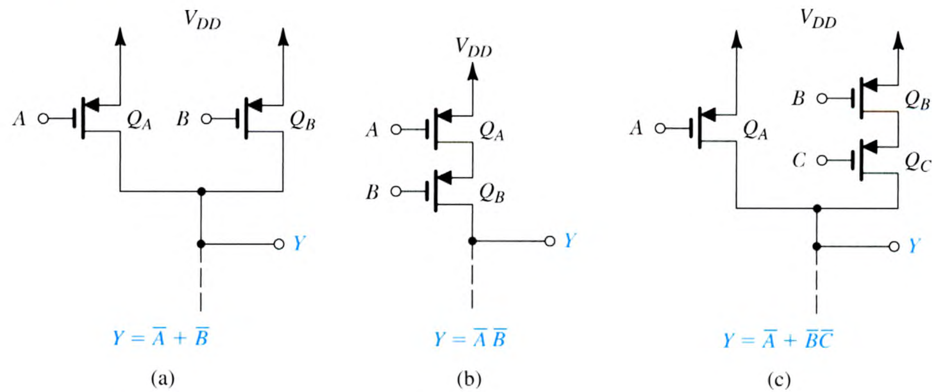


Figure 16.5 Examples of pull-up networks.

the inputs are high. In a dual manner, the PUN comprises PMOS transistors, and a PMOS transistor conducts when the input signal at its gate is low; thus the PUN is activated when the inputs are low.

The PDN utilizes devices in parallel to form an OR function and devices in series to form an AND function; the same is true of the PUN. Here, the OR and AND notations refer to current flow or conduction. Figure 16.4 shows examples of PDNs. For the circuit in Fig. 16.4(a), Q_A will conduct when A is high ($v_A = V_{DD}$) and will then pull the output node down to ground ($v_Y = 0$ V, $Y = 0$). Similarly, Q_B conducts and pulls Y down when B is high. Thus Y will be low when A is high *or* B is high, which can be expressed as

$$\overline{Y} = A + B$$

or equivalently

$$Y = \overline{A + B}$$

The PDN in Fig. 16.4(b) will conduct only when A and B are both high simultaneously. Thus Y will be low when A is high *and* B is high,

$$\overline{Y} = AB$$

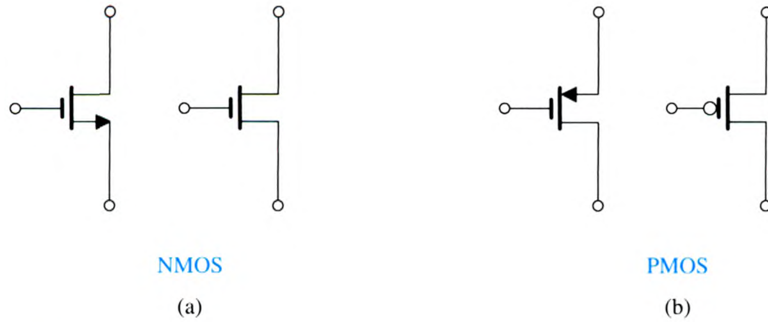


Figure 16.6 Usual and alternative circuit symbols for MOSFETs.

or equivalently

$$Y = \overline{AB}$$

As a final example, the PDN in Fig. 16.4(c) will conduct and cause Y to be 0 when A is high *or* when B *and* C are both high, thus

$$\overline{Y} = A + BC$$

or equivalently

$$Y = \overline{A + BC}$$

Next consider the PUN examples shown in Fig. 16.5. The PUN in Fig. 16.5(a) will conduct and pull Y up to V_{DD} ($Y = 1$) when A is low *or* B is low, thus

$$Y = \overline{A} + \overline{B}$$

The PUN in Fig. 16.5(b) will conduct and produce a high output ($v_Y = V_{DD}$, $Y = 1$) only when A *and* B are both low, thus

$$Y = \overline{A}\,\overline{B}$$

Finally, the PUN in Fig. 16.5(c) will conduct and cause Y to be high (logic 1) if A is low *or* if B *and* C are both low; thus,

$$Y = \overline{A} + \overline{B}\,\overline{C}$$

Having developed an understanding and appreciation of the structure and operation of PDNs and PUNs, we are almost ready to consider complete CMOS gates. First, however, we need to get to know the alternative circuit symbols that digital-circuit designers almost universally use for MOS transistors. Figure 16.6 shows our usual symbols (left) and the corresponding "digital" symbols (right). Notice that the symbol for the PMOS transistor with a circle at the gate terminal is intended to indicate that the signal at the gate has to be low for the device to be activated (i.e., to conduct). Thus, in terms of logic-circuit terminology, the gate terminal of the PMOS transistor is an *active low* input. Besides indicating this property of PMOS devices, the digital symbols omit any indication of which of the device terminals is the source and which is the drain. This should cause no difficulty at this stage of our study; simply remember that for an NMOS transistor, the drain is

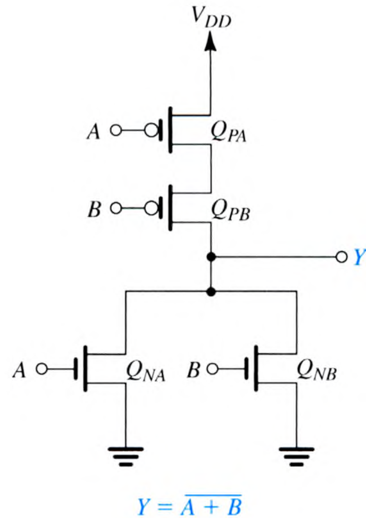


Figure 16.7 A two-input CMOS NOR gate.

the terminal at the higher voltage (current flows from drain to source), and for a PMOS transistor the source is the terminal at the higher voltage (current flows from source to drain). To be consistent with the literature, we will now use these modified symbols for MOS transistors in logic applications, except where our usual symbols help in understanding circuit operation.

16.1.4 The Two-Input NOR Gate

We first consider the CMOS gate that realizes the two-input NOR function

$$Y = \overline{A + B} = \overline{A}\,\overline{B} \tag{16.1}$$

We can see that Y is to be low (PDN conducting) when A is high or B is high. Thus the PDN consists of two parallel NMOS devices with A and B as inputs [i.e., the circuit in Fig. 16.4(a)]. For the PUN, we note from the second expression in Eq. (16.1) that Y is to be high when A and B are both low. Thus the PUN consists of two series PMOS devices with A and B as the inputs [i.e., the circuit in Fig. 16.5(b)]. Putting the PDN and the PUN together gives the CMOS NOR gate shown in Fig. 16.7. Note that extension to a higher number of inputs is straightforward: For each additional input, an NMOS transistor is added in parallel with Q_{NA} and Q_{NB}, and a PMOS transistor is added in series with Q_{PA} and Q_{PB}.

16.1.5 The Two-Input NAND Gate

The two-input NAND function is described by the Boolean expression

$$Y = \overline{AB} = \overline{A} + \overline{B} \tag{16.2}$$

To synthesize the PDN, we consider the input combinations that require Y to be low: There is only one such combination, namely, A and B both high. Thus, the PDN simply comprises two NMOS transistors in series [such as the circuit in Fig. 16.4(b)]. To synthesize the PUN, we consider the input combinations that result in Y being high. These are found from the second expression in Eq. (16.2) as A low or B low. Thus, the PUN consists of two parallel PMOS

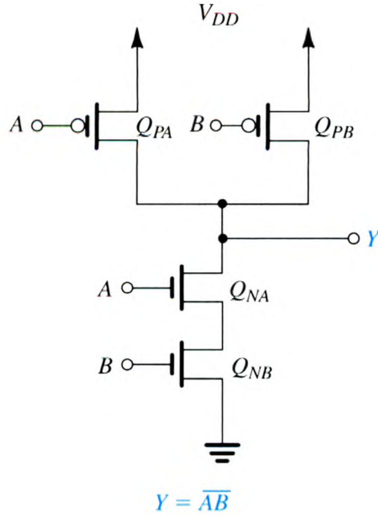


Figure 16.8 A two-input CMOS NAND gate.

transistors with A and B applied to their gates [such as the circuit in Fig. 16.5(a)]. Putting the PDN and PUN together results in the CMOS NAND gate implementation shown in Fig. 16.8. Note that extension to a higher number of inputs is straightforward: For each additional input, we add an NMOS transistor in series with Q_{NA} and Q_{NB}, and a PMOS transistor in parallel with Q_{PA} and Q_{PB}.

16.1.6 A Complex Gate

Consider next the more complex logic function

$$Y = \overline{A(B+CD)} \tag{16.3}$$

Since $\overline{Y} = A(B+CD)$, we see that Y should be low for A high and simultaneously either B high or C and D both high, from which the PDN is directly obtained. To obtain the PUN, we need to express Y in terms of the complemented variables. We do this through repeated application of DeMorgan's law, as follows:

$$\begin{aligned} Y &= \overline{A(B+CD)} \\ &= \overline{A} + \overline{B+CD} \\ &= \overline{A} + \overline{B}\,\overline{CD} \\ &= \overline{A} + \overline{B}(\overline{C}+\overline{D}) \end{aligned} \tag{16.4}$$

Thus, Y is high for A low or B low and either C or D low. The corresponding complete CMOS circuit will be as shown in Fig. 16.9.

16.1.7 Obtaining the PUN from the PDN and Vice Versa

From the CMOS gate circuits considered thus far (e.g., that in Fig. 16.9), we observe that the PDN and the PUN are dual networks: Where a series branch exists in one, a parallel branch exists in the other. Thus, we can obtain one from the other, a process that can be simpler

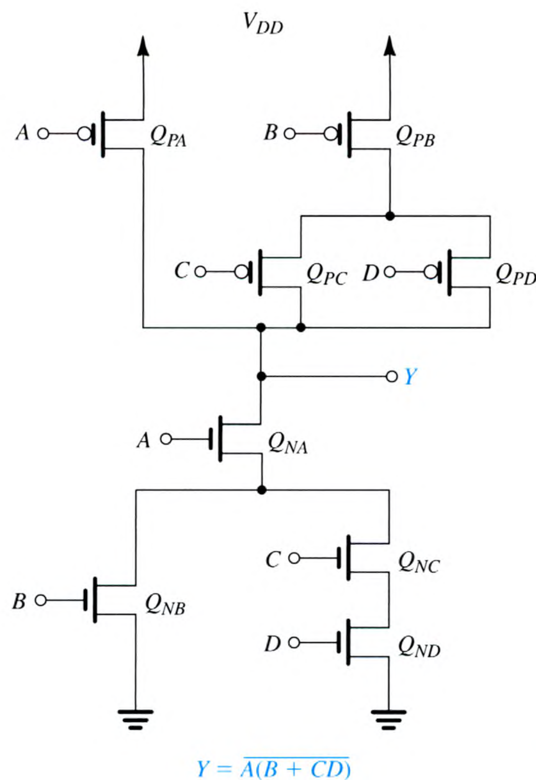


Figure 16.9 CMOS realization of a complex gate.

than having to synthesize each separately from the Boolean expression of the function. For instance, in the circuit of Fig. 16.9, we found it relatively easy to obtain the PDN, simply because we already had $\overline{Y}$ in terms of the uncomplemented inputs. On the other hand, to obtain the PUN, we had to manipulate the given Boolean expression to express Y as a function of the complemented variables, the form convenient for synthesizing PUNs. Alternatively, we could have used this duality property to obtain the PUN from the PDN. The reader is urged to refer to Fig. 16.9 to convince herself that this is indeed possible.

We should, however, mention that at times it is not easy to obtain one of the two networks from the other using the duality property. For such cases, we would have to resort to a more rigorous process, which is beyond the scope of this book (see Kang, Leblebici, and Kim, 2016).

16.1.8 The Exclusive-OR Function

An important function that often arises in logic design is the exclusive-OR (XOR) function,

$$Y = A\overline{B} + \overline{A}B \tag{16.5}$$

We observe that since Y (rather than $\overline{Y}$) is given, it is easier to synthesize the PUN. Unfortunately, however, Y is not a function of the complemented variables only (as we would like it to be). Thus, we will need additional inverters. The PUN obtained directly from Eq. (16.5) is shown in Fig. 16.10(a). Note that the Q_1, Q_2 branch realizes the first term

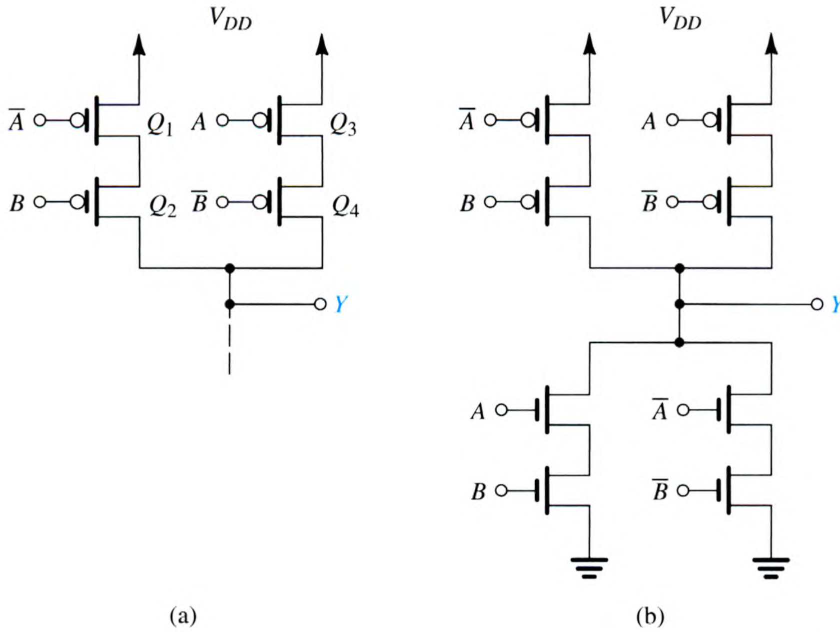


Figure 16.10 Realization of the exclusive-OR (XOR) function. **(a)** The PUN synthesized directly from the expression in Eq. (16.5). **(b)** The complete XOR realization utilizing the PUN in **(a)** and a PDN that is synthesized directly from the expression in Eq. (16.6). Note that two inverters (not shown) are needed to generate the complemented variables. Also note that in this XOR realization, the PDN and the PUN are not dual networks; however, a realization based on dual networks is possible (see Problem 16.9).

($A\overline{B}$), whereas the Q_3, Q_4 branch realizes the second term ($\overline{A}B$). Note also the need for two additional inverters to generate $\overline{A}$ and $\overline{B}$.

As for synthesizing the PDN, we can obtain it as the dual network of the PUN in Fig. 16.10(a). Alternatively, we can develop an expression for $\overline{Y}$ and use it to synthesize the PDN. Leaving the first approach for the reader to do as an exercise, we shall utilize the direct synthesis approach. DeMorgan's law can be applied to the expression in Eq. (16.5) to obtain $\overline{Y}$ as

$$\overline{Y} = AB + \overline{A}\,\overline{B} \tag{16.6}$$

The corresponding PDN will be as in Fig. 16.10(b), which shows the CMOS realization of the exclusive-OR function except for the two additional inverters. Note that the exclusive-OR requires 12 transistors for its realization, a rather complex network.

Another interesting observation follows from the circuit in Fig. 16.10(b). The PDN and the PUN here are *not* dual networks. Indeed, duality of the PDN and the PUN is not a necessary condition. Thus, although a dual of PDN (or PUN) can always be used for PUN (or PDN), the two networks are not necessarily duals.

16.1.9 Summary of the Synthesis Method

1. The PDN can be most directly synthesized by expressing $\overline{Y}$ as a function of the *uncomplemented* variables. If complemented variables appear in this expression, additional inverters will be required to generate them.

2. The PUN can be most directly synthesized by expressing Y as a function of the *complemented* variables and then applying the uncomplemented variables to the gates of the PMOS transistors. If uncomplemented variables appear in the expression, additional inverters will be needed.
3. The PDN can be obtained from the PUN (and vice versa) using the duality property.

Example 16.1

Synthesize a CMOS logic circuit that implements the Boolean function

$$Y = \overline{A + B(C + D)}$$

Solution

To obtain the PDN we use

$$\overline{Y} = A + B(C + D)$$

and note that Y will be low when A is high or when $[B(C + D)]$ is high. Thus we have two parallel networks. One consists of a single transistor with A supplied to its gate, and the second is a network composed of a transistor controlled by B in series with two parallel transistors controlled by C and D. The resulting PDN is shown in Fig. 16.11(a).

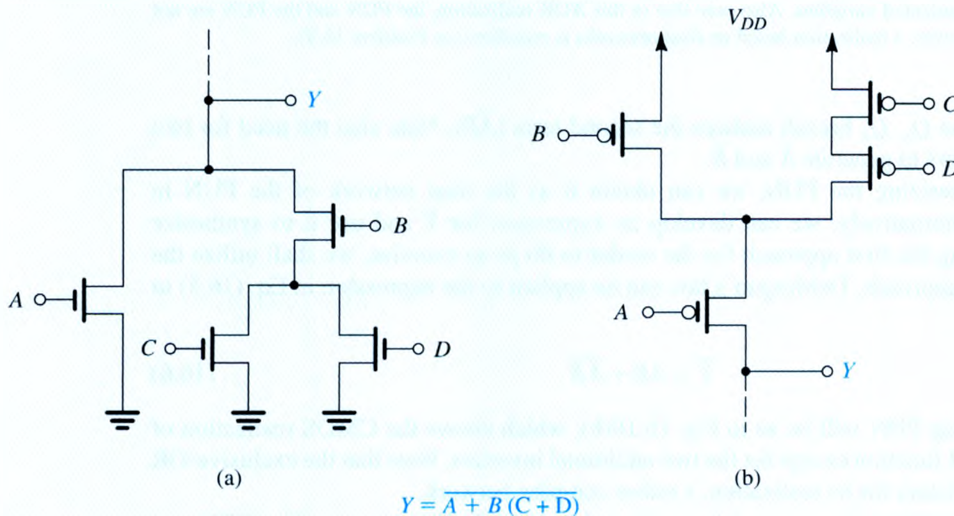


Figure 16.11 **(a)** The PDN and **(b)** the PUN for the logic function in Example 16.1.

A question arises: Should the B transistor be placed closer to ground (with the C and D transistors closer to Y) or closer to Y? From a logic standpoint, *both* are valid solutions.

To obtain the PUN we use deMorgan's law on the given expression for Y to obtain

$$Y = \overline{A}(\overline{B} + \overline{C}\,\overline{D})$$

which leads to the circuit in Fig. 16.11(b). Here again, we may exchange the positions of the C and D transistors with no effect on the logic behavior. Finally, the complete circuit can be obtained by combining the PDN and the PUN shown in Fig. 16.11.

EXERCISE

16.1 What logic function does the following PUN realize?
Ans. $Y = \overline{A}\,\overline{B} + \overline{C}\,\overline{D}\,\overline{E}$

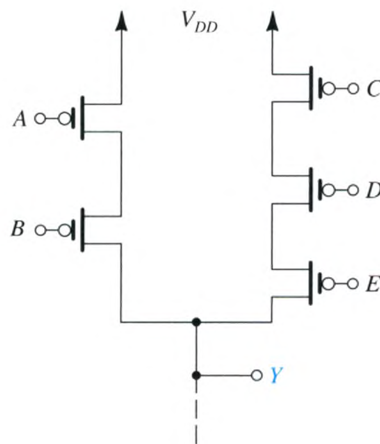


Figure E16.1

16.2 Digital Logic Inverters

Having learned how to synthesize CMOS circuits that implement various logic functions, we next consider the design and performance evaluation of these CMOS logic circuits. Toward that end, we step back to study in detail the characterization of the basic logic element, the inverter. We will do this in this section in general terms; that is, our study will not be confined to CMOS inverters only. In subsequent sections we specialize what we will have learned in this section to the case of the CMOS inverter and extend it to CMOS logic gates.

16.2.1 The Voltage-Transfer Characteristic (VTC)

Refer to the inverter shown in block form in Fig. 16.1(a). To quantify the operation of the inverter, we utilize its voltage-transfer characteristic (VTC). We have already introduced the concept of the VTC and utilized it to characterize the operation of basic MOSFET amplifiers

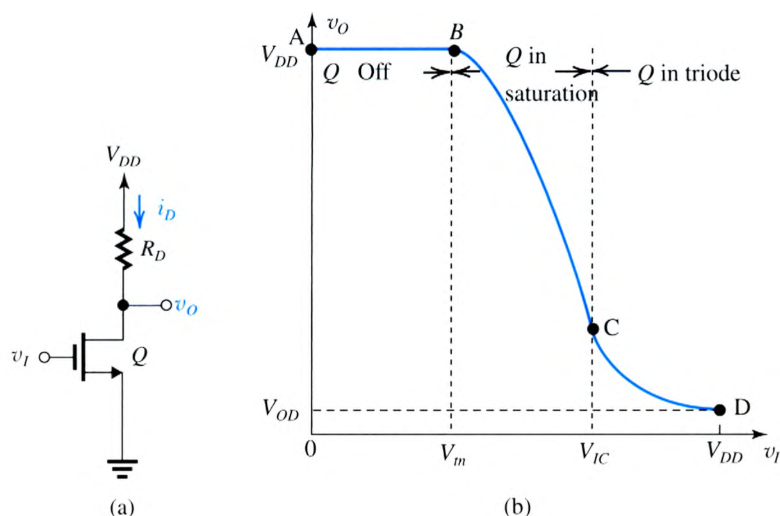


Figure 16.12 The simple resistively loaded MOS amplifier can be used as a logic inverter when operated in cutoff $(v_I < V_{tn})$ and in triode $(v_I > V_{IC})$. The output high level is V_{DD} and the low level is V_{OD}.

in Section 7.1.3. Figure 16.12 shows such a circuit, together with its VTC. Observe that the circuit in fact implements the inverter function: For a logic-0 input, v_I is close to 0 V and specifically lower than the MOSFET threshold voltage V_{tn}, the transistor will be off, $i_D = 0$, and $v_O = V_{DD}$, which is a logic 1. For a logic-1 input, $v_I = V_{DD}$, the transistor will be conducting and operating in the triode region (at point D on the VTC), and the output voltage will be low (logic 0).

Thus to use this amplifier as a logic inverter, we utilize its extreme regions of operation. This is exactly the opposite to its use as a signal amplifier, where it would be biased at the middle of the transfer characteristic segment BC and the signal kept small enough to restrict operation to a short, almost linear, segment of the transfer curve. Digital applications, on the other hand, make use of the gross nonlinearity exhibited by the VTC.

With these observations in mind, we show in Fig. 16.13 a possible VTC of a logic inverter. For simplicity, we are using three straight lines to approximate the VTC, which is usually a nonlinear curve such as that in Fig. 16.12. Observe that the output high level, denoted V_{OH}, does not depend on the exact value of v_I as long as v_I does not exceed the value labeled V_{IL}; when v_I exceeds V_{IL}, the output decreases and the inverter enters its amplifier region of operation, also called the **transition region**. It follows that V_{IL} is an important parameter of the inverter VTC: It is the *maximum value that v_I can have while being interpreted by the inverter as representing a logic 0.*

Similarly, we observe that the output low level, denoted V_{OL}, does not depend on the exact value of v_I as long as v_I does not fall below V_{IH}. Thus V_{IH} is an important parameter of the inverter VTC: It is the *minimum value that v_I can have while being interpreted by the inverter as representing a logic 1.*

16.2.2 Noise Margins

The insensitivity of the inverter output to the exact value of v_I within allowed regions is a great advantage that digital circuits have over analog circuits. To quantify this insensitivity property, consider the situation that occurs often in a digital system where an inverter (or

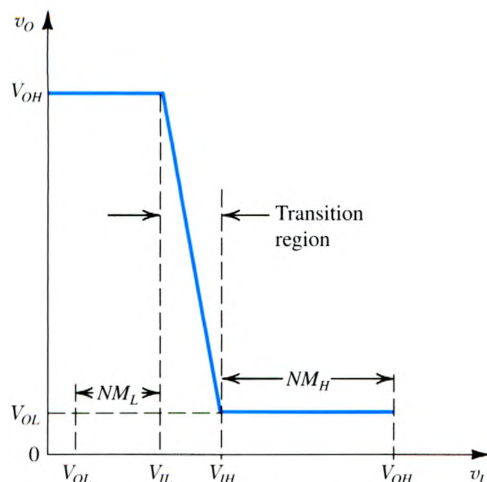


Figure 16.13 Voltage-transfer characteristic of an inverter. The VTC is approximated by three straight-line segments. Note the four parameters of the VTC (V_{OH}, V_{OL}, V_{IL}, and V_{IH}) and their use in determining the noise margins (NM_H and NM_L).

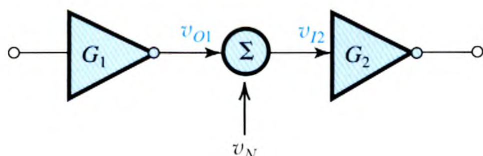


Figure 16.14 Noise voltage v_N is coupled to the interconnection between the output of inverter G_1 and the input of inverter G_2.

a logic gate based on the inverter circuit) is driving another similar inverter, as shown in Fig. 16.14.

Here we assume that a noise or interference signal v_N is somehow coupled to the interconnection between the output of inverter G_1 and the input of inverter G_2 with the result that the input of G_2 becomes

$$v_{I2} = v_{O1} + v_N \tag{16.7}$$

where the noise voltage v_N can be either positive or negative. Now consider the case $v_{O1} = V_{OL}$; that is, inverter G_2 is driven by a logic-0 signal. Reference to Fig. 16.13 indicates that in this case G_2 will continue to function properly as long as its input v_{I2} does not exceed V_{IL}. Equation (16.7) then indicates that v_N can be as high as $V_{IL} - V_{OL}$ while G_2 continues to function properly. Thus, we can say that inverter G_2 has a **noise margin for low input**, NM_L, of

$$NM_L = V_{IL} - V_{OL} \tag{16.8}$$

Similarly, if $v_{O1} = V_{OH}$, the driven inverter G_2 will continue to see a high input as long as v_{I2} does not fall below V_{IH}. Thus, in the high-input state, inverter G_2 can tolerate a negative v_N of magnitude as high as $V_{OH} - V_{IH}$. We can thus state that G_2 has a **high-input noise margin**, NM_H, of

$$NM_H = V_{OH} - V_{IH} \tag{16.9}$$

In summary, four parameters, V_{OH}, V_{OL}, V_{IH}, and V_{IL}, define the VTC of an inverter and determine its noise margins, which in turn measure the ability of the inverter to tolerate variations in the input signal levels. In this regard, observe that changes in the input signal level within the noise margins are *rejected* by the inverter. Thus noise is not allowed to propagate further through the system, a definite advantage of digital over analog circuits. Alternatively,

Table 16.1 Important Parameters of the VTC of the Logic Inverter (Refer to Fig. 16.13)

V_{OL}:	Output low level
V_{OH}:	Output high level
V_{IL}:	Maximum value of input interpreted by the inverter as a logic 0
V_{IH}:	Minimum value of input interpreted by the inverter as a logic 1
NM_L:	Noise margin for low input $= V_{IL} - V_{OL}$
NM_H:	Noise margin for high input $= V_{OH} - V_{IH}$

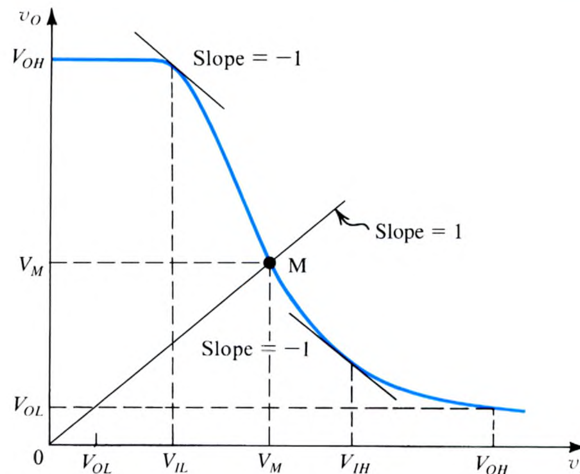


Figure 16.15 Typical voltage-transfer characteristic (VTC) of a logic inverter, illustrating the definition of the critical points.

we can think of the inverter as *restoring* the signal levels to standard values (V_{OL} and V_{OH}) even when it is presented with corrupted input signal levels (within the noise margins). As a summary, useful for future reference, we present a listing and definitions of the important parameters of the inverter VTC in Table 16.1. For easy reference, Table 16.1 is also presented in the Summary Tables supplement found online at www.oup.com/he/sedra-smith8e.

The formal definitions of the threshold voltages V_{IL} and V_{IH} are given in Fig. 16.15. Observe that V_{IL} and V_{IH} are defined as the VTC points at which the slope is -1 V/V. As v_I exceeds V_{IL}, the magnitude of the inverter gain increases and the VTC enters its transition region. Similarly, as v_I falls below V_{IH}, the inverter enters the transition region and the magnitude of the gain increases. Finally, note that Fig. 16.15 shows the definition of another important point on the VTC; this is point M at which $v_O = v_I$. Point M is loosely considered to be the midpoint of the VTC and thus the point at which the *inverter switches from one state to the other*. Point M plays an important role in the definition of the time delay of the inverter, as we shall see in Chapter 17.

16.2.3 The Ideal VTC

What constitutes an ideal VTC for an inverter? The answer to this naturally arising question follows directly from the preceding discussion: An ideal VTC is one that maximizes the output signal swing and the noise margins. For an inverter operated from a power supply V_{DD}, maximum signal swing is obtained when

$$V_{OH} = V_{DD}$$

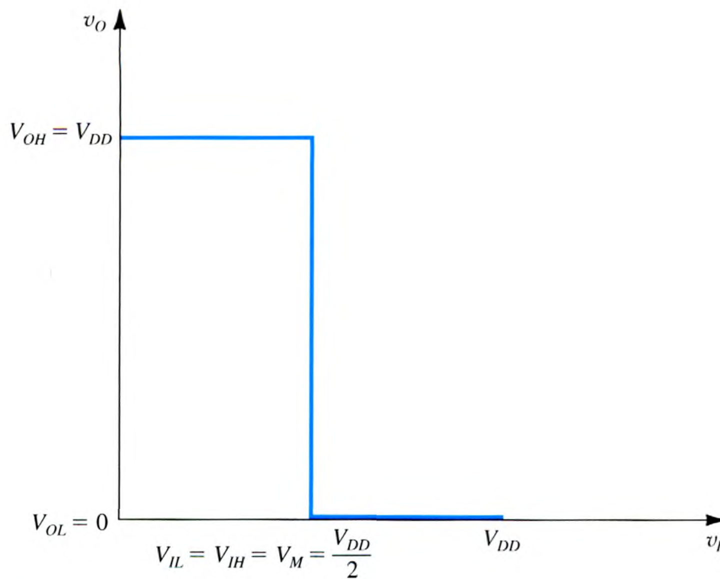


Figure 16.16 The VTC of an ideal inverter.

and

$$V_{OL} = 0$$

To obtain maximum noise margins, we first arrange for the transition region to be made as narrow as possible and ideally of zero width. Then, the two noise margins are equalized by arranging for the transition from high to low to occur at the midpoint of the power supply, that is, at $V_{DD}/2$. The result is the VTC shown in Fig. 16.16, for which

$$V_{IL} = V_{IH} = V_M = V_{DD}/2$$

Observe that the sharp transition at $V_{DD}/2$ indicates that if the inverter were to be used as an amplifier, its gain would be infinite. Again, we point out that while the analog designer's interest would be focused on the transition region of the VTC, the digital designer would prefer the transition region to be as narrow as possible, as is the case in the ideal VTC of Fig. 16.16. Finally, we will see in Section 16.3 that inverters implemented using CMOS technology come very close to realizing the ideal VTC.

16.2.4 Inverter Implementation

Inverters are implemented using transistors (Chapters 5 and 6) operating as **voltage-controlled switches**. The simplest inverter implementation is shown in Fig. 16.17(a). The switch is controlled by the inverter input voltage v_I: When v_I is low, the switch will be open and $v_O = V_{DD}$, since no current flows through R. When v_I is high, the switch will be closed and, assuming an ideal switch, v_O will be 0.

Transistor switches, however, as we know from Chapters 5 and 6, are not perfect. Although their **off-resistances** are very high, and thus an open switch closely approximates an open circuit, the "on" switch has a finite closure, or **on-resistance**, R_{on}. The result is that

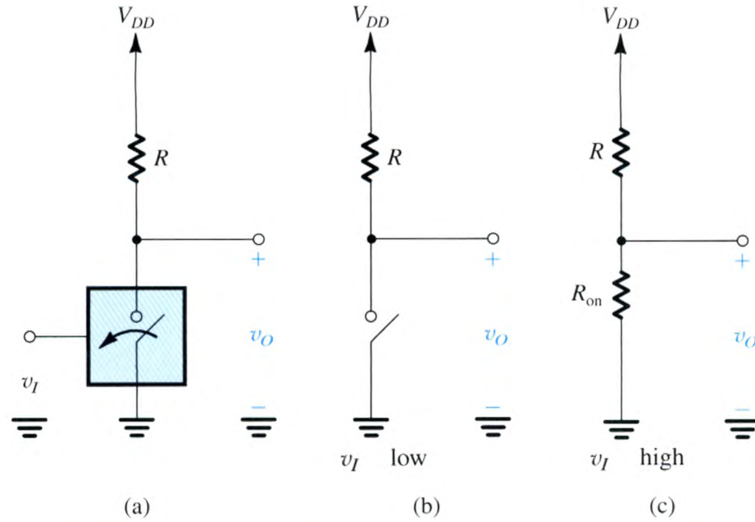


Figure 16.17 **(a)** The simplest implementation of a logic inverter using a voltage-controlled switch; **(b)** equivalent circuit when v_I is low; **(c)** equivalent circuit when v_I is high. Note that the switch is assumed to close when v_I is high.

when v_I is high, the inverter has the equivalent circuit shown in Fig. 16.17(c), from which V_{OL} can be found.

$$V_{OL} = V_{DD}\frac{R_{on}}{R + R_{on}}$$

We observe that the circuit in Fig. 16.12(a) is a direct implementation of the inverter in Fig. 16.17. In this case, R_{on} is equal to r_{DS} of the MOSFET evaluated at its operating point in the triode region with $V_{GS} = V_{DD}$.

EXERCISE

D16.2 Design the inverter in Fig. 16.12(a) to provide $V_{OL} = 40$ mV and to draw a supply current of 20 μA in the low-output state. Let the transistor be specified to have $V_t = 0.35$ V, $\mu_n C_{ox} = 540\ \mu\text{A/V}^2$, and $\lambda = 0$. The power supply $V_{DD} = 1.0$ V. Specify the required values of W/L and R_D. How much power is drawn from V_{DD} when the switch is open? Closed?

Hint: Recall that for small v_{DS},

$$r_{DS} \simeq 1\Big/\left[(\mu_n C_{ox})\left(\frac{W}{L}\right)(V_{GS} - V_t)\right]$$

Ans. 1.42; 48 kΩ; 0; 20 μW

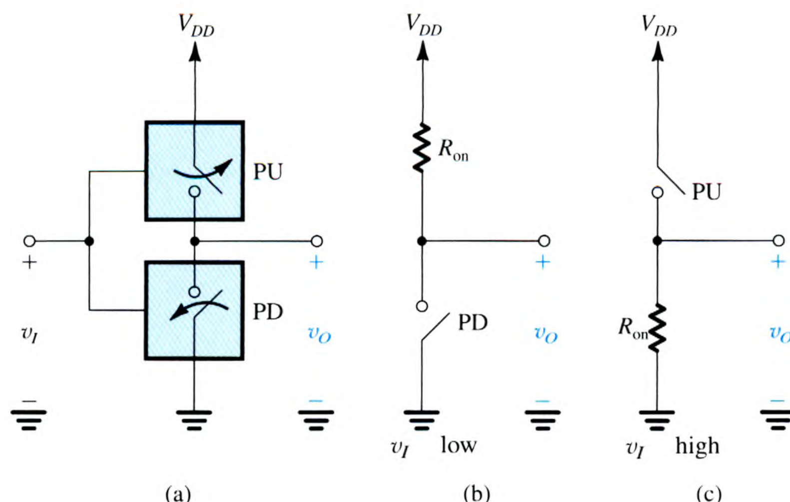


Figure 16.18 A more elaborate implementation of the logic inverter utilizing two complementary switches. This is the basis of the CMOS inverter that we synthesized in the previous section [Fig. 16.2(b)] and shall study in Section 16.3.

More elaborate implementations of the logic inverter exist, and we show two of these in Fig. 16.18(a) and 16.19. The circuit in Fig. 16.18(a) utilizes a pair of **complementary switches**; the **"pull-up" (PU) switch** connects the output node to V_{DD}, and the **"pull-down" (PD) switch** connects the output node to ground. When v_I is low, the PU switch will be closed and the PD switch open, resulting in the equivalent circuit of Fig. 16.18(b). Observe that in this case R_{on} of PU connects the output to V_{DD}, thus establishing $V_{OH} = V_{DD}$. Also observe that no current flows, and thus no power is dissipated, in the circuit. Next, if v_I is raised to the logic-1 level, the PU switch will open while the PD switch will close, resulting in the equivalent circuit shown in Fig. 16.18(c). Here R_{on} of the PD switch connects the output to ground, thus establishing $V_{OL} = 0$. Here again no current flows, and no power is dissipated. The superiority of this inverter implementation over that using the single pull-down switch and a resistor (known as a **pull-up resistor**) should be obvious: With $V_{OL} = 0$ and $V_{OH} = V_{DD}$, the signal swing is at its maximum possible, and the power dissipation is zero in both states. This circuit constitutes the basis of the CMOS inverter that we synthesized in the previous section [Fig. 16.2(b)] and will study in detail in Section 16.3.

Next, consider the inverter implementation of Fig. 16.19. Here a double-throw switch is used to steer the constant current I_{EE} into one of two resistors connected to the positive supply V_{CC}. The reader is urged to show that if a high v_I results in the switch being connected to R_{C1}, then a logic inversion function is realized at v_{O1}. Note that the output voltage is independent of the switch resistance. Also, there is a structural similarity between the *current-steering* or *current-mode* logic arrangement and the differential pair circuits described in Chapter 8.

Resistively Loaded MOS Inverter The inverter first shown in Fig 16.12(a), which is composed of a MOSFET that acts as a switch and a resistor R_D that acts as a load, is known as a resistively loaded MOS inverter. Example 16.2 investigates the output voltage levels and noise margins of this type of inverter, and demonstrates that V_{OL}, large R_D, and power dissipation can become limiting factors in the inverter's performance.

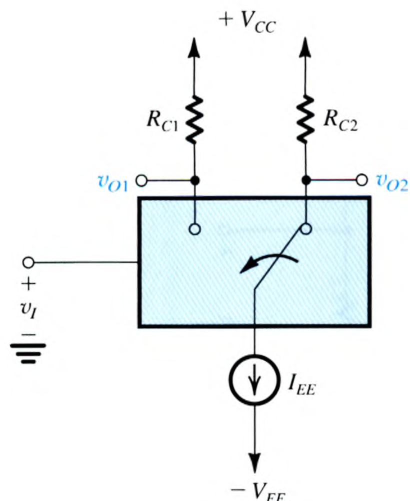


Figure 16.19 Another inverter implementation utilizing a double-throw switch to steer the constant current I_{EE} to R_{C1} (when v_I is high) or R_{C2} (when v_I is low). This is the basis of a family of logic circuits called emitter-coupled logic (ECL).

Example 16.2 Resistively Loaded MOS Inverter

For the simple MOS inverter in Fig. 16.12(a):

(a) Derive expressions for V_{OH}, V_{OL}, V_{IL}, V_{IH}, and V_M. For simplicity, neglect channel-length modulation (i.e., assume $\lambda = 0$). Show that these inverter parameters can be expressed in terms of V_{DD}, V_t, and $(k_n R_D)$. The latter parameter has the dimension of V^{-1}, and to simplify the expressions, denote $k_n R_D \equiv 1/V_x$.

(b) Show that V_x can be used as a design parameter for the inverter circuit. In particular, find the value of V_x that results in $V_M = V_{DD}/2$.

(c) Find numerical values for all parameters and for the inverter noise margins for $V_{DD} = 1.0$ V, $V_t = 0.35$ V, and V_x set to the value found in (b).

(d) For $k_n' = 540\ \mu\text{A/V}^2$ and $W/L = 1.5$, find the required value of R_D and use it to determine the average power dissipated in the inverter, assuming that the inverter spends half of the time in each of its two states.

(e) Comment on the characteristics of this inverter circuit vis-à-vis the ideal characteristics as well as on its suitability for implementation in integrated-circuit form.

Solution

(a) Refer to Fig. 16.20. For $v_I < V_t$, the MOSFET is off, $i_D = 0$, and $v_O = V_{DD}$. Thus

$$V_{OH} = V_{DD} \tag{16.10}$$

As v_I exceeds V_t, the MOSFET turns on and operates initially in the saturation region. Assuming $\lambda = 0$,

$$i_D = \frac{1}{2}k_n(v_I - V_t)^2$$

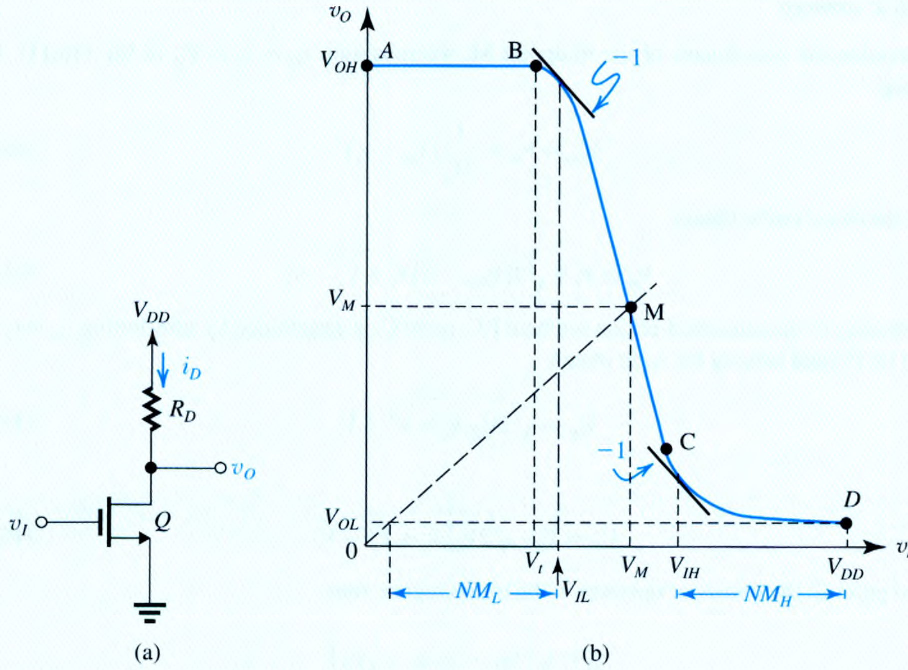


Figure 16.20 The resistively loaded MOS inverter and its VTC (Example 16.2).

and

$$v_O = V_{DD} - R_D i_D = V_{DD} - \frac{1}{2} k_n R_D (v_I - V_t)^2$$

substituting $k_n R_D = 1/V_x$, the BC segment of the VTC is described by

$$v_O = V_{DD} - \frac{1}{2V_x}(v_I - V_t)^2 \qquad (16.11)$$

To determine V_{IL}, we differentiate Eq. (16.11) and set $dv_O/dv_I = -1$,

$$\frac{dv_O}{dv_I} = -\frac{1}{V_x}(v_I - V_t)$$

$$-1 = -\frac{1}{V_x}(V_{IL} - V_t)$$

which results in

$$V_{IL} = V_t + V_x \qquad (16.12)$$

Example 16.2 *continued*

To determine the coordinates of the midpoint M, we substitute $v_O = v_I = V_M$ in Eq. (16.11), thus obtaining

$$V_{DD} - V_M = \frac{1}{2V_x}(V_M - V_t)^2 \tag{16.13}$$

which can be solved to obtain

$$V_M = V_t + \sqrt{2(V_{DD} - V_t)V_x + V_x^2} - V_x \tag{16.14}$$

The boundary of the saturation-region segment BC, point C, is determined by substituting $v_O = v_I - V_t$ in Eq. (16.11) and solving for v_O to obtain

$$V_{OC} = \sqrt{2V_{DD}V_x + V_x^2} - V_x \tag{16.15}$$

and

$$V_{IC} = V_t + \sqrt{2V_{DD}V_x + V_x^2} - V_x \tag{16.16}$$

Beyond point C, the transistor operates in the triode region, thus

$$i_D = k_n\left[(v_I - V_t)v_O - \frac{1}{2}v_O^2\right]$$

and the output voltage is obtained as

$$v_O = V_{DD} - \frac{1}{V_x}\left[(v_I - V_t)v_O - \frac{1}{2}v_O^2\right] \tag{16.17}$$

which describes the segment CD of the VTC. To determine V_{IH}, we differentiate Eq. (16.17) and set $dv_O/dv_I = -1$:

$$\frac{dv_O}{dv_I} = -\left(\frac{1}{V_x}\right)\left[(v_I - V_t)\frac{dv_O}{dv_I} + v_O - v_O\frac{dv_O}{dv_I}\right]$$

$$-1 = -\frac{1}{V_x}[-(V_{IH} - V_t) + 2v_O]$$

which results in

$$V_{IH} - V_t = 2v_O - V_x \tag{16.18}$$

Substituting in Eq. (16.17) for v_I with the value of V_{IH} from Eq. (16.18) results in an equation in the value of v_O corresponding to $v_I = V_{IH}$, which can be solved to yield

$$v_O\big|_{v_I = V_{IH}} = 0.816\sqrt{V_{DD}V_x} \tag{16.19}$$

which can be substituted in Eq. (16.18) to obtain

$$V_{IH} = V_t + 1.63\sqrt{V_{DD}V_x} - V_x \tag{16.20}$$

To determine V_{OL} we substitute $v_I = V_{OH} = V_{DD}$ in Eq. (16.17):

$$V_{OL} = V_{DD} - \frac{1}{V_x}\left[(V_{DD} - V_t)V_{OL} - \frac{1}{2}V_{OL}^2\right] \tag{16.21}$$

Since we expect V_{OL} to be much smaller than $2(V_{DD} - V_t)$, we can approximate Eq. (16.21) as

$$V_{OL} \simeq V_{DD} - \frac{1}{V_x}(V_{DD} - V_t)V_{OL}$$

which results in

$$V_{OL} = \frac{V_{DD}}{1 + [(V_{DD} - V_t)/V_x]} \tag{16.22}$$

It is interesting to note that the value of V_{OL} can alternatively be found by noting that at point D, the MOSFET switch has a closure resistance r_{DS},

$$r_{DS} = \frac{1}{k_n(V_{DD} - V_t)} \tag{16.23}$$

and V_{OL} can be obtained from the voltage divider formed by R_D and r_{DS},

$$V_{OL} = V_{DD}\frac{r_{DS}}{R_D + r_{DS}} = \frac{V_{DD}}{1 + R_D/r_{DS}} \tag{16.24}$$

Substituting for r_{DS} from Eq. (16.23) gives an expression for V_{OL} identical to that in Eq. (16.22).

(b) We observe that all the inverter parameters derived above are functions of V_{DD}, V_t, and V_x only. Since V_{DD} and V_t are determined by the process technology, the only design parameter available is $V_x \equiv 1/k_nR_D$. To place V_M at half the supply voltage V_{DD}, we substitute $V_M = V_{DD}/2$ in Eq. (16.13) to obtain the value V_x must have as

$$V_x\big|_{V_M = V_{DD}/2} = \frac{(V_{DD}/2 - V_t)^2}{V_{DD}} \tag{16.25}$$

(c) For $V_{DD} = 1.0\text{V}$ and $V_t = 0.35$, we use Eq. (16.25) to obtain

$$V_x\big|_{V_M = 0.5\text{ V}} = \frac{(1.0/2 - 0.35)^2}{1.0} = 0.0225\text{ V}$$

From Eq. (16.10): $V_{OH} = 1.0$ V

From Eq. (16.22): $V_{OL} = 33$ mV

From Eq. (16.12): $V_{IL} = 0.37$ V

From Eq. (16.20): $V_{IH} = 0.57$ V

Example 16.2 *continued*

$$NM_L = V_{IL} - V_{OL} = 0.34 \text{ V}$$
$$NM_H = V_{OH} - V_{IH} = 0.43 \text{ V}$$

(d) To determine R_D, we use

$$k_n R_D = \frac{1}{V_x} = \frac{1}{0.0225} = 44.4$$

Thus,

$$R_D = \frac{44.4}{k_n'(W/L)} = \frac{44.4}{540 \times 10^{-6} \times 1.5} = 54.9 \text{ k}\Omega$$

The inverter dissipates power only when the output is low, in which case the current drawn from the supply is

$$I_{DD} = \frac{V_{DD} - V_{OL}}{R_D} = \frac{1.0 - 0.033}{54.9 \text{ k}\Omega} = 17.6 \ \mu\text{A}$$

and the power drawn from the supply during the low-output interval is

$$P_D = V_{DD} I_{DD} = 1.0 \times 17.6 = 17.6 \ \mu\text{W}$$

Since the inverter spends half of the time in this state,

$$P_{D\text{average}} = \frac{1}{2} P_D = 8.8 \ \mu\text{W}$$

(e) We now can make a few comments on the characteristics of this inverter circuit in comparison to the ideal characteristics:

1. The output signal swing, though not equal to the full power supply, is reasonably good: $V_{OH} = 1.0$ V, $V_{OL} = 0.033$ V.
2. The noise margins, though of reasonable values, are far from the optimum value of $V_{DD}/2$. This is particularly the case for NM_L.
3. Most seriously, the gate dissipates a relatively large amount of power. To appreciate this point, consider an IC chip with a million inverters (a small number by today's standards): Its power dissipation will be 9 W. This is too large, especially given that this is "static power," unrelated to the switching activity of the gates (more on this later).

We consider this inverter implementation to be entirely unsuitable for IC fabrication because each inverter requires a load resistance of 54.9 kΩ, a value that needs a large chip area (see Appendix A). To overcome this problem, we can replace the passive resistance R_D with a PMOS transistor, a technique we call "pseudo-NMOS logic."

EXERCISES

D16.3 In an attempt to reduce the required value of R_D to 10 kΩ, the designer of the inverter in Example 16.2 decides to keep the parameter V_x unchanged but increases W/L. What is the new value required for W/L? Do the noise margins change? What does the power dissipation become?
Ans. 8.22; no; 48 μW

D16.4 In an attempt to reduce the required value of R_D to 10 kΩ, the designer of the inverter in Example 16.2 decides to change V_x while keeping W/L unchanged. What new value of V_x is needed? What do the noise margins become? What does the power dissipation become?
Ans. $V_x = 0.12$ V; $NM_L = 0.31$ V, $NM_H = 0.20$ V; 42 μW

Pseudo-NMOS Inverter To eliminate the problem associated with the need for a large resistance R_D in the resistively loaded inverter, R_D can be replaced by a MOSFET. One such possibility is the circuit in Fig. 16.21, where the load is a PMOS transistor Q_P whose gate is tied to ground in order to turn it on. Because of its resemblance to an earlier form of logic (NMOS logic, now obsolete) in which the load is an NMOS transistor, this circuit is known as a pseudo-NMOS inverter.

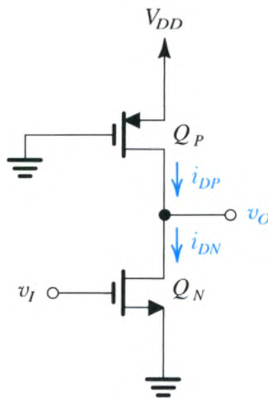


Figure 16.21 Pseudo-NMOS inverter.

16.3 The CMOS Inverter

In this section we study the inverter circuit of the most widely used digital IC technology: CMOS. The basic CMOS inverter, synthesized in Section 16.1.2, is shown in Fig. 16.22. It utilizes two MOSFETs: one, Q_N, with an n channel and the other, Q_P, with a p channel. The body of each device is connected to its source, and thus no body effect arises. As will be seen shortly, the CMOS circuit realizes the conceptual inverter implementation studied in the previous section (Fig. 16.18), where a pair of switches are operated in a complementary fashion by the input voltage v_I.

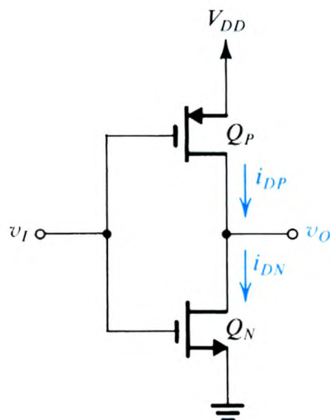


Figure 16.22 The CMOS inverter.

16.3.1 Circuit Operation

We first consider the two extreme cases: when v_I is at logic-0 level, which is 0 V, and when v_I is at logic-1 level, which is V_{DD} volts. In both cases, for ease of exposition we shall consider the n-channel device Q_N to be the driving transistor and the p-channel device Q_P to be the load. However, since the circuit is symmetric, this assumption is obviously arbitrary, and the reverse would lead to identical results.

Figure 16.23 illustrates the case when $v_I = V_{DD}$, showing the $i_D - v_{DS}$ characteristic curve for Q_N with $v_{GSN} = V_{DD}$. (Note that $i_D = i$ and $v_{DSN} = v_O$.) Superimposed on the Q_N characteristic curve is the load curve, which is the $i_D - v_{SD}$ curve of Q_P for the case $v_{SGP} = 0$ V. Since $v_{SGP} < |V_t|$, the load curve will be a horizontal straight line at zero current level. The operating point will be at the intersection of the two curves, where we note that the output voltage is zero and the current through the two devices is also zero. This means that the power dissipation in the circuit is zero. Note, however, that although Q_N is operating at zero current and zero drain-source voltage (i.e., at the origin of the $i_D - v_{DS}$ plane), the operating point is on a steep segment of the $i_D - v_{DS}$ characteristic curve. Thus Q_N provides a low-resistance path between the output terminal and ground, with the resistance obtained using Eq. (5.13b) as

$$r_{DSN} = 1 \Big/ \left[k_n' \left(\frac{W}{L} \right)_n (V_{DD} - V_{tn}) \right] \tag{16.26}$$

Figure 16.23(c) shows the equivalent circuit of the inverter when the input is high.[1] This circuit confirms that $v_O \equiv V_{OL} = 0$ V and that the power dissipation in the inverter is zero.

The other extreme case, when $v_I = 0$ V, is illustrated in Fig. 16.24. In this case Q_N is operating at $v_{GSN} = 0$; hence its $i_D - v_{DS}$ characteristic is a horizontal straight line at zero current level. The load curve is the $i_D - v_{SD}$ characteristic of the p-channel device with $v_{SGP} = V_{DD}$. As shown, at the operating point the output voltage is equal to V_{DD}, and the current in the two devices is still zero. Thus the power dissipation in the circuit is zero in both extreme states.

[1] In Section 16.1 we referred to r_{DSN} (and r_{DSP} for p-channel devices) as R_{on}.

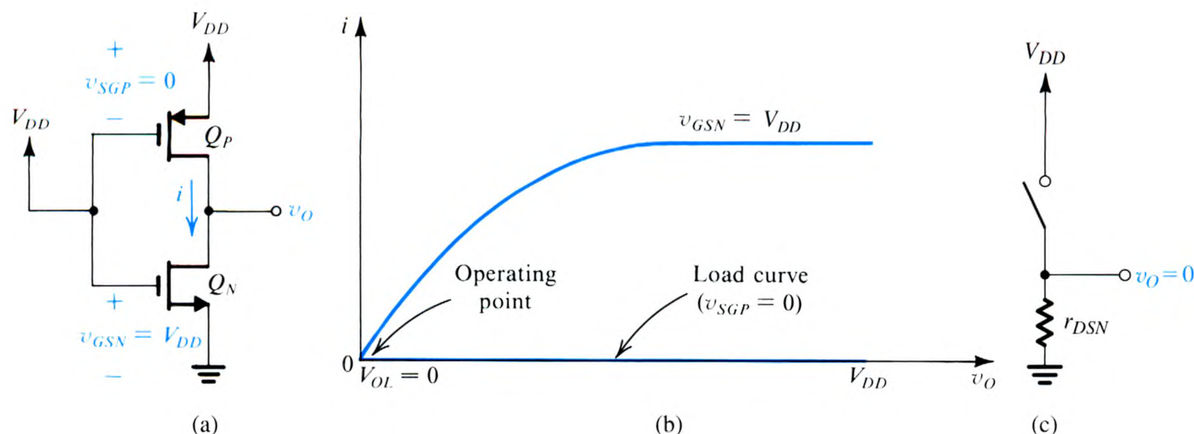


Figure 16.23 Operation of the CMOS inverter when v_I is high: **(a)** circuit with $v_I = V_{DD}$ (logic-1 level, or V_{OH}); **(b)** graphical construction to determine the operating point; **(c)** equivalent circuit.

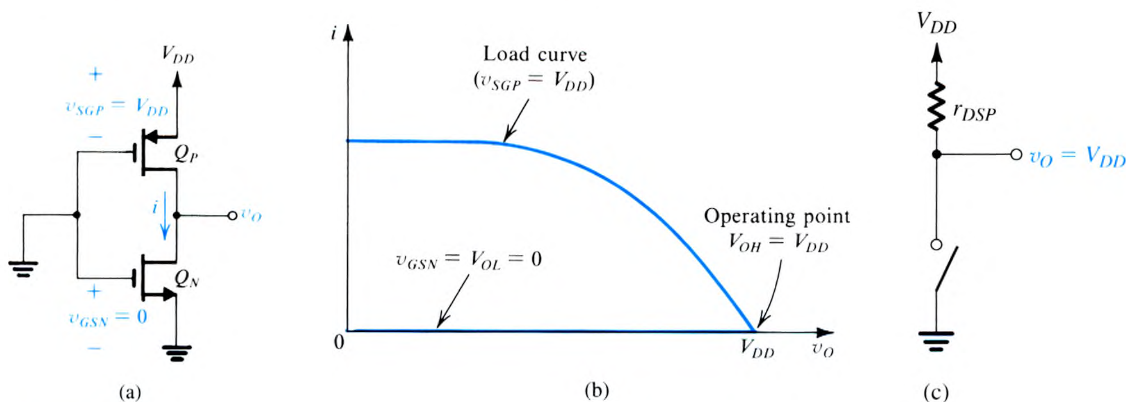


Figure 16.24 Operation of the CMOS inverter when v_I is low: **(a)** circuit with $v_I = 0$ V (logic-0 level, or V_{OL}); **(b)** graphical construction to determine the operating point; **(c)** equivalent circuit.

Figure 16.24(c) shows the equivalent circuit of the inverter when the input is low. Here we see that Q_P provides a low-resistance path between the output terminal and the dc supply V_{DD}, with the resistance given by

$$r_{DSP} = 1 \Big/ \left[k'_p \left(\frac{W}{L} \right)_p (V_{DD} - |V_{tp}|) \right] \tag{16.27}$$

The equivalent circuit confirms that in this case $v_O \equiv V_{OH} = V_{DD}$ and that the power dissipation in the inverter is zero.

It should be noted, however, that in spite of the fact that the quiescent current is zero, the load-driving capability of the CMOS inverter is high. For instance, with the input high, as in the circuit of Fig. 16.23, transistor Q_N can sink a relatively large load current. This current can quickly discharge the load capacitance, as will be seen in Chapter 17. Because of its action in sinking load current and thus pulling the output voltage down toward ground, transistor Q_N is known as the pull-down device. Similarly, with the input low, as in the circuit of Fig. 16.24, transistor Q_P can source a relatively large load current. This current can quickly charge up a

load capacitance, thus pulling the output voltage up toward V_{DD}. Hence, Q_P is known as the pull-up device. The reader will recall that we used this terminology in connection with the conceptual inverter circuit of Fig. 16.18 and in Section 16.1 as well.

From the above, we conclude that the basic CMOS logic inverter behaves as an ideal inverter. In summary:

1. The output voltage levels are 0 and V_{DD}, and thus the signal swing is the maximum possible. This, coupled with the fact that the inverter can be designed to provide a symmetrical voltage-transfer characteristic, results in wide noise margins.
2. The static power dissipation in the inverter is zero (neglecting the dissipation due to leakage currents) in both of its states. This is because no dc path exists between the power supply and ground in either state.
3. A low-resistance path exists between the output terminal and ground (in the low-output state) or V_{DD} (in the high-output state). These low-resistance paths ensure that the output voltage is 0 or V_{DD} independent of the exact values of the W/L ratios or other device parameters. Furthermore, the low output resistance makes the inverter less sensitive to the effects of noise and other disturbances.
4. The active pull-up and pull-down devices provide the inverter with high output-driving capability in both directions. As will be seen in Section 17.1, this speeds up the operation considerably.
5. The input resistance of the inverter is infinite (because $I_G = 0$). Thus the inverter can drive an arbitrarily large number of similar inverters with no loss in signal level. Of course, each additional inverter increases the load capacitance on the driving inverter and slows down the operation. In Section 17.1, we will consider the inverter switching times.

FRANK MARION WANLESS: THE INVENTOR OF CMOS

While working for Fairchild Semiconductor in 1963, Frank Wanless filed the first patent on CMOS logic, heralding the new age of zero-static-power logic. In 1964, as director of research and engineering at General Microelectronics (a start-up later bought by Philco-Ford), he created the first commercial CMOS integrated circuit. The symmetry of the logic form Wanless had invented was at first emphasized by the use of the name COmplementary Symmetry MOS, or COS-MOS, but the simpler CMOS shorthand soon prevailed.

16.3.2 The Voltage-Transfer Characteristic (VTC)

The complete voltage-transfer characteristic (VTC) of the CMOS inverter can be obtained by repeating the graphical procedure, used above in the two extreme cases, for all intermediate values of v_I. In the following, we shall calculate the critical points of the resulting voltage-transfer curve. For this we need the i–v relationships of Q_N and Q_P. For Q_N,

$$i_{DN} = k_n'\left(\frac{W}{L}\right)_n\left[(v_I - V_{tn})v_O - \frac{1}{2}v_O^2\right] \quad \text{for } v_O \le v_I - V_{tn} \tag{16.28}$$

and

$$i_{DN} = \frac{1}{2}k_n'\left(\frac{W}{L}\right)_n(v_I - V_{tn})^2 \quad \text{for } v_O \ge v_I - V_{tn} \tag{16.29}$$

For Q_P,

$$i_{DP} = k_p'\left(\frac{W}{L}\right)_p\left[(V_{DD} - v_I - |V_{tp}|)(V_{DD} - v_O) - \frac{1}{2}(V_{DD} - v_O)^2\right]$$
$$\text{for } v_O \geq v_I + |V_{tp}| \qquad (16.30)$$

and

$$i_{DP} = \frac{1}{2}k_p'\left(\frac{W}{L}\right)_p(V_{DD} - v_I - |V_{tp}|)^2 \qquad \text{for } v_O \leq v_I + |V_{tp}| \qquad (16.31)$$

The CMOS inverter is usually designed to have $V_{tn} = |V_{tp}| = V_t$. Also, although this is not always the case, we shall assume that Q_N and Q_P are matched; that is, $k_n'(W/L)_n = k_p'(W/L)_p$. It should be noted that since μ_p is often 0.25 to 0.5 times the value of μ_n, to make $k'(W/L)$ of the two devices equal, the width of the p-channel device is made two to four times that of the n-channel device. More specifically, the two devices are designed to have equal lengths, with widths related by

$$\frac{W_p}{W_n} = \frac{\mu_n}{\mu_p} \qquad (16.32)$$

This will result in $k_n'(W/L)_n = k_p'(W/L)_p$, and the inverter will have a symmetric transfer characteristic and equal current-driving capability in both directions (pull-up and pull-down).

With Q_N and Q_P matched, the CMOS inverter has the voltage-transfer characteristic shown in Fig. 16.25. As indicated, the transfer characteristic has five distinct segments corresponding to different combinations of modes of operation of Q_N and Q_P. The vertical segment BC is obtained when both Q_N and Q_P are operating in the saturation region. Because we are neglecting the finite output resistance in saturation, that is, assuming $\lambda_N = \lambda_P = 0$, the inverter gain in this region is infinite. From symmetry, this vertical segment occurs at $v_I = V_{DD}/2$ and is bounded by $v_O(\text{B}) = V_{DD}/2 + V_t$, at which value Q_P enters the triode region and $v_O(\text{C}) = V_{DD}/2 - V_t$, at which value Q_N enters the triode region.

The reader will recall from Section 16.2.1 that in addition to V_{OL} and V_{OH}, two other points on the transfer curve determine the noise margins of the inverter. These are the maximum permitted logic-0 or "low" level at the input, V_{IL}, and the minimum permitted logic-1 or "high" level at the input, V_{IH}. These are formally defined as the two points on the transfer curve at which the incremental gain is unity (i.e., the slope is – 1 V/V).

To determine V_{IH}, we note that Q_N is in the triode region, and thus its current is given by Eq. (16.28), while Q_P is in saturation and its current is given by Eq. (16.31). Equating i_{DN} and i_{DP}, and assuming matched devices, gives

$$(v_I - V_t)v_O - \frac{1}{2}v_O^2 = \frac{1}{2}(V_{DD} - v_I - V_t)^2 \qquad (16.33)$$

Differentiating both sides relative to v_I results in

$$(v_I - V_t)\frac{dv_O}{dv_I} + v_O - v_O\frac{dv_O}{dv_I} = -(V_{DD} - v_I - V_t)$$

in which we substitute $v_I = V_{IH}$ and $dv_O/dv_I = -1$ to obtain

$$v_O = V_{IH} - \frac{V_{DD}}{2} \qquad (16.34)$$

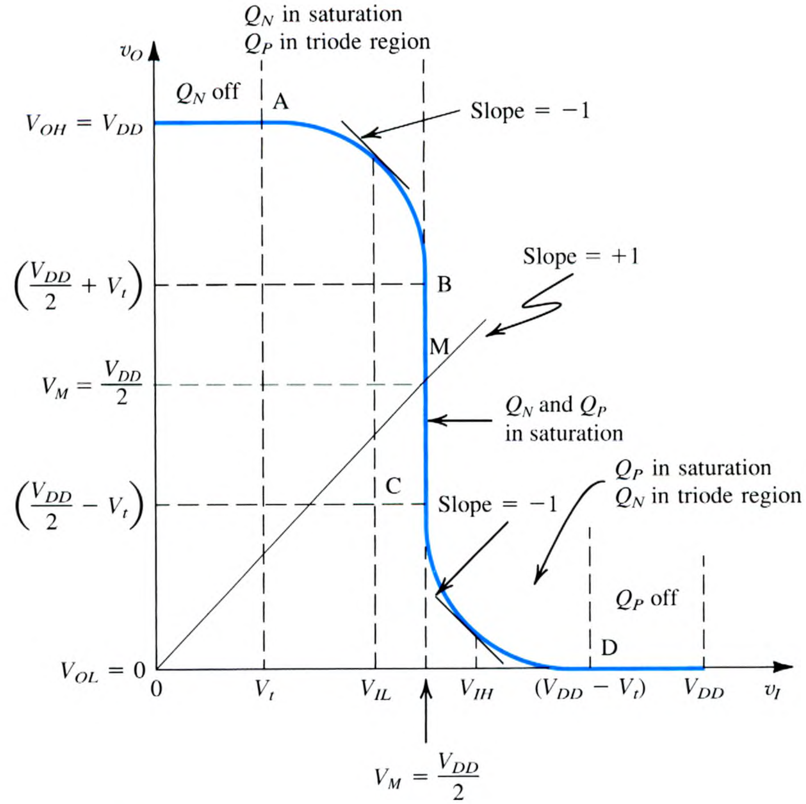


Figure 16.25 The voltage-transfer characteristic of the CMOS inverter when Q_N and Q_P are matched.

Substituting $v_I = V_{IH}$ and for v_O from Eq. (16.34) in Eq. (16.33) gives

$$V_{IH} = \frac{1}{8}(5V_{DD} - 2V_t) \tag{16.35}$$

V_{IL} can be determined in a manner similar to that used to find V_{IH}. Alternatively, we can use the symmetry relationship

$$V_{IH} - \frac{V_{DD}}{2} = \frac{V_{DD}}{2} - V_{IL}$$

together with V_{IH} from Eq. (16.35) to obtain

$$V_{IL} = \frac{1}{8}(3V_{DD} + 2V_t) \tag{16.36}$$

The noise margins can now be determined as follows:

$$\begin{aligned} NM_H &= V_{OH} - V_{IH} \\ &= V_{DD} - \frac{1}{8}(5V_{DD} - 2V_t) \\ &= \frac{1}{8}(3V_{DD} + 2V_t) \end{aligned} \tag{16.37}$$

$$
\begin{aligned}
NM_L &= V_{IL} - V_{OL} \\
&= \frac{1}{8}(3V_{DD} + 2V_t) - 0 \\
&= \frac{1}{8}(3V_{DD} + 2V_t)
\end{aligned}
\tag{16.38}
$$

As expected, the symmetry of the voltage-transfer characteristic results in equal noise margins. Of course, if Q_N and Q_P are not matched, the voltage-transfer characteristic will no longer be symmetric, and the noise margins will not be equal.

16.3.3 The Situation When Q_N and Q_P Are Not Matched

In the above, we assumed that Q_N and Q_P are matched; that is, in addition to $V_{tn} = |V_{tp}|$, the transconductance parameters k_n and k_p are made equal by selecting W_p/W_n according to Eq. (16.32). The result is a symmetrical VTC that switches at the midpoint of the supply; that is, $V_M = V_{DD}/2$. The symmetry, as we have seen, equalizes and maximizes the noise margins.

The price paid for obtaining a perfectly symmetric VTC is that the width of the p-channel device can be three to four times as large as that of the n-channel device. This can result in a relatively large silicon area, which, besides being wasteful of silicon real estate, can also result in increased device capacitances and a corresponding increase in the propagation delay of the inverter (Section 17.1). It is useful, therefore, to inquire into the effect of not matching Q_N and Q_P. Toward that end we derive an expression for the switching voltage V_M as follows.

Since at M, both Q_N and Q_P operate in saturation, their currents are given by Eqs. (16.29) and (16.31), respectively. Substituting $v_I = v_O = V_M$, and equating the two currents results in

$$V_M = \frac{r(V_{DD} - |V_{tp}|) + V_{tn}}{r + 1} \tag{16.39}$$

where

$$r = \sqrt{\frac{k_p}{k_n}} = \sqrt{\frac{\mu_p W_p}{\mu_n W_n}} \tag{16.40}$$

where we have assumed that Q_N and Q_P have the same channel length L, which is usually the case with L equal to the minimum available for the given process technology. Note that the matched case corresponds to $r = 1$. For $|V_{tp}| = V_{tn}$, and $r = 1$, Eq. (16.39) yields $V_M = V_{DD}/2$, as expected. For a given process, that is, given values for V_{DD}, V_{tn}, and V_{tp}, one can plot V_M versus the matching parameter r. Such a plot, for a 65-nm process, is shown in Fig. 16.26. We make the following two observations:

1. V_M increases with r. Thus, making $k_p > k_n$ shifts V_M toward V_{DD}. Conversely, making $k_p < k_n$ shifts V_M toward 0.
2. V_M is not a strong function of r. For the particular case shown, lowering r by a factor of 2 (from 1 to 0.5), reduces V_M by only 0.05 V.

Observation 2 implies that if one is willing to tolerate a small reduction in NM_L, substantial savings in silicon area can be obtained. This point is illustrated in Example 16.3.

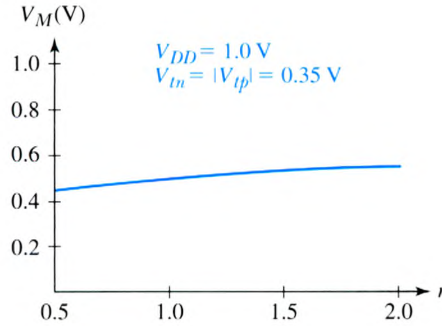


Figure 16.26 Variation of the inverter switching voltage, V_M, with the parameter $r = \sqrt{k_p/k_n}$.

Example 16.3 CMOS Inverter Static Characteristics and Design

Consider a CMOS inverter fabricated in a 65-nm process for which $V_{DD} = 1.0$ V, $V_{tn} = |V_{tp}| = 0.35$ V, $\mu_n = 5.4\mu_p$, and $\mu_n C_{ox} = 540\ \mu\text{A/V}^2$. In addition, Q_N and Q_P have $L = 65$ nm and $(W/L)_n = 1.5$.

(a) Find W_p that results in $V_M = V_{DD}/2 = 0.5$ V. What is the silicon area utilized by the inverter in this case?

(b) For the matched case in (a), find the values of V_{OH}, V_{OL}, V_{IH}, V_{IL}, and the noise margins NM_L and NM_H. For $v_I = V_{IH}$, what value of v_O results? This can be considered the worst-case value of V_{OL}. Similarly, for $v_I = V_{IL}$, find v_O that is the worst-case value of V_{OH}. Now, use these worst-case values to determine more conservative values for the noise margins.

(c) For the matched case in (a), find the output resistance of the inverter in each of its two states.

(d) If $\lambda_n = |\lambda_p| = 5\ \text{V}^{-1}$, what is the inverter gain at $v_I = V_M$? If a straight line is drawn through the point $v_I = v_O = V_M$ with a slope equal to the gain, at what values of v_I does it intercept the horizontal lines $v_O = 0$ and $v_O = V_{DD}$? Use these intercepts to estimate the width of the transition region of the VTC.

(e) If $W_p = W_n$, what value of V_M results? What do you estimate the reduction of NM_L (relative to the matched case) to be? What is the percentage savings in silicon area (relative to the matched case)?

(f) Repeat (e) for the case $W_p = 2W_n$. This case, which is frequently used in industry, can be considered to be a compromise between the minimum-area case in (e) and the matched case.

Solution

(a) To obtain $V_M = V_{DD}/2 = 0.5$ V, we select W_p according to Eq. (16.32),

$$\frac{W_p}{W_n} = \frac{\mu_n}{\mu_p} = 5.4$$

Since $W_n/L = 1.5$, $W_n = 1.5 \times 65 = 97.5$ nm.[2] Thus,

$$W_p = 5.4 \times 97.5 = 526.5 \text{ nm}$$

[2]In these technologies, we note that transistor widths are often selected to be multiples of a unit width, and would not typically be expressed as a fraction of a nanometer.

For this design, the silicon area is

$$A = W_n L + W_p L = L(W_n + W_p)$$
$$= 65(97.5 + 526.5) = 40{,}560 \text{ nm}^2$$

(b)

$$V_{OH} = V_{DD} = 1.0 \text{ V}$$
$$V_{OL} = 0 \text{ V}$$

To obtain V_{IH} we use Eq. (16.35),

$$V_{IH} = \frac{1}{8}(5V_{DD} - 2V_t) = \frac{1}{8}(5 \times 1.0 - 2 \times 0.35) = 0.54 \text{ V}$$

To obtain V_{IL} we use Eq. (16.36),

$$V_{IL} = \frac{1}{8}(3V_{DD} + 2V_t) = \frac{1}{8}(3 \times 1.0 + 2 \times 0.35) = 0.46 \text{ V}$$

We can now compute the noise margins as

$$NM_H = V_{OH} - V_{IH} = 1.0 - 0.54 = 0.46 \text{ V}$$
$$NM_L = V_{IL} - V_{OL} = 0.46 - 0 = 0.46 \text{ V}$$

As expected, $NM_H = NM_L$, and their value is very close to the optimum value of $V_{DD}/2 = 0.5$ V.

For $v_I = V_{IH} = 0.54$ V, we can obtain the corresponding value of v_O by substituting in Eq. (16.34),

$$v_O = V_{IH} - \frac{V_{DD}}{2} = 0.54 - \frac{1.0}{2} = 0.04 \text{ V}$$

Thus, the worst-case value of V_{OL}, that is, $V_{OL\max}$, is 0.04 V, and the noise margin NM_L reduces to

$$NM_L = V_{IL} - V_{OL\max} = 0.46 - 0.04 = 0.42 \text{ V}$$

From symmetry, we can obtain the value of v_O corresponding to $v_I = V_{IL}$ as

$$v_O = V_{DD} - 0.04 = 0.96 \text{ V}$$

Thus the worst-case value of V_{OH}, that is, $V_{OH\min}$, is 0.96 V, and the noise margin NM_H reduces to

$$NM_H = V_{OH\min} - V_{IH} = 0.96 - 0.54 = 0.42 \text{ V}$$

Note that the reduction in the noise margins is slight.

Example 16.3 *continued*

(c) The output resistance of the inverter in the low-output state is

$$r_{DSN} = \frac{1}{\mu_n C_{ox}(W/L)_n(V_{DD} - V_{tn})}$$
$$= \frac{1}{540 \times 10^{-6} \times 1.5(1.0 - 0.35)} = 1.9 \text{ k}\Omega$$

Since Q_N and Q_P are matched, the output resistance in the high-output state will be equal, that is,

$$r_{DSP} = r_{DSN} = 1.9 \text{ k}\Omega$$

(d) If the inverter is biased to operate at $v_I = v_O = V_M = 0.5$ V, then each of Q_N and Q_P will be operating at an overdrive voltage $V_{OV} = V_M - V_t = 0.5 - 0.35 = 0.15$ V and will be conducting equal dc currents I_D of

$$I_D = \frac{1}{2}\mu_n C_{ox}\left(\frac{W}{L}\right)_N V_{OV}^2$$
$$= \frac{1}{2} \times 540 \times 1.5 \times 0.15^2$$
$$= 9.1 \ \mu\text{A}$$

Thus, Q_N and Q_P will have equal transconductances:

$$g_{mn} = g_{mp} = \frac{2I_D}{V_{OV}} = \frac{2 \times 9.1}{0.15} = 0.12 \text{ mA/V}$$

Transistors Q_N and Q_P will have equal output resistances r_o,

$$r_{on} = r_{op} = \frac{|V_A|}{I_D} = \frac{1}{|\lambda| I_D} = \frac{1}{5 \times 9.1} = 21.9 \text{ k}\Omega$$

We can now compute the voltage gain at M as

$$A_v = -(g_{mn} + g_{mp})(r_{on} \parallel r_{op})$$
$$= -(0.12 + 0.12)(21.9 \parallel 21.9) = -2.6 \text{ V/V}$$

When the straight line at M of slope −2.6 V/V is extrapolated, it intersects the line $v_O = 0$ at $[0.5 + 0.5/2.6] = 0.69$ V and the line $v_O = V_{DD}$ at $0.5 - 0.5/2.6 = 0.31$ V. Thus the width of the transition region can be considered to be $(0.69 - 0.31) = 0.38$ V.

(e) For $W_p = W_n$, the parameter r can be found from Eq. (16.40),

$$r = \sqrt{\frac{\mu_p W_p}{\mu_n W_n}} = \sqrt{\frac{1}{5.4} \times 1} = 0.43$$

The corresponding value of V_M can be determined from Eq. (16.39) as

$$V_M = \frac{0.43(1.0 - 0.35) + 0.35}{0.43 + 1} = 0.44 \text{ V}$$

Thus V_M shifts by only -0.06 V. Without recalculating V_{IL} we can estimate the reduction in NM_L to be approximately equal to the shift in V_M, that is, NM_L becomes $0.46 - 0.06 = 0.40$ V. The silicon area for this design can be computed as follows:

$$A = L(W_n + W_p) = 65(97.5 + 97.5)$$
$$= 12{,}675 \text{ nm}^2$$

This represents a 69% reduction from the matched case!

(f) For $W_p = 2W_n$,

$$r = \sqrt{\frac{1}{5.4} \times 2} = 0.609$$

$$V_M = \frac{0.609(1.0 - 0.35) + 0.35}{0.609 + 1} = 0.46 \text{ V}$$

Thus, relative to the matched case, the shift in V_M is only -0.04 V. We estimate that NM_L will decrease from 0.46 V by the same amount; thus NM_L becomes 0.42 V. In this case, the silicon area required is

$$A = L(W_n + W_p) = 65(97.5 + 195)$$
$$= 19{,}012.5 \text{ nm}^2$$

which represents a 53% reduction relative to the matched case!

EXERCISES

16.5 Consider a CMOS inverter fabricated in a 0.13-μm process for which $V_{DD} = 1.2$ V, $V_{tn} = -V_{tp} = 0.4$ V, $\mu_n/\mu_p = 4$, and $\mu_n C_{ox} = 430\ \mu\text{A/V}^2$. In addition, Q_N and Q_P have $L = 0.13$ μm and $(W/L)_n = 1.0$.

(a) Find W_p that results in $V_M = 0.6$ V.

(b) For the matched case in (a), find the values of V_{OH}, V_{OL}, V_{IH}, V_{IL}, NM_H, and NM_L.

(c) For the inverter in (a), find the output resistance in each of its two states.

(d) For a minimum-size inverter for which $(W/L)_p = (W/L)_n = 1.0$, find V_M.

Ans. (a) 0.52 μm; (b) 1.2 V, 0 V, 0.65 V, 0.55 V, 0.55 V, 0.55 V; (c) 2.9 kΩ, 2.9 kΩ; (d) 0.53 V

D16.6 A CMOS inverter utilizes $V_{DD} = 5$ V, $V_{tn} = |V_{tp}| = 1$ V, and $\mu_n C_{ox} = 2\mu_p C_{ox} = 50\ \mu\text{A/V}^2$. Find $(W/L)_n$ and $(W/L)_p$ so that $V_M = 2.5$ V and so that for $v_I = V_{DD}$, the inverter can sink a current of 0.2 mA with the output voltage not exceeding 0.2 V.

Ans. $(W/L)_n \simeq 5$; $(W/L)_p \simeq 10$

Summary

- A CMOS logic gate consists of an NMOS pull-down network (PDN) and a PMOS pull-up network (PUN). The PDN conducts for every input combination that requires a low output. Since an NMOS transistor conducts when its input is high, the PDN is most directly synthesized from the expression for the low output $(\overline{Y})$ as a function of the uncomplemented inputs. In a complementary fashion, the PUN conducts for every input combination that corresponds to a high output. Since a PMOS conducts when its input is low, the PUN is most directly synthesized from the expression for a high output (Y) as a function of the complemented inputs.
- The digital logic inverter is the basic building block of digital circuits, just as the amplifier is the basic building block of analog circuits.
- The static operation of a logic inverter is described by its voltage-transfer characteristic (VTC). The VTC determines the inverter noise margins; refer to Fig. 16.13, Fig. 16.15, and Table 16.1 for the definitions of important VTC points and the noise margins. In particular, note that $NM_H = V_{OH} - V_{IH}$ and $NM_L = V_{IL} - V_{OL}$, and refer to the ideal VTC in Fig. 16.16.
- The inverter is implemented using transistors operating as voltage-controlled switches. There are three possible arrangements, shown in Figs. 16.17, 16.18, and 16.19. The arrangement in Fig. 16.18 results in a high-performance inverter and is the basis for the CMOS inverter studied in Section 16.3.
- Table 16.2 provides a summary of the important characteristics of the CMOS inverter. For easy reference, this table is also presented as Table 16.2 in the Summary Tables supplement found online at www.oup.com/he/sedra-smith8e.

Table 16.2 Summary of Important Static Characteristics of the CMOS Logic Inverter

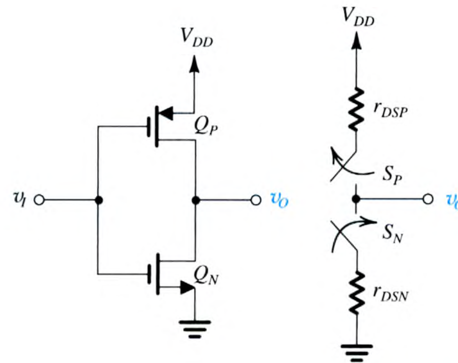
Inverter Output Resistance

- When v_O is low (current sinking):

$$r_{DSN} = 1\Big/\left[k_n'\left(\frac{W}{L}\right)_n(V_{DD}-V_{tn})\right]$$

- When v_O is high (current sourcing):

$$r_{DSP} = 1\Big/\left[k_p'\left(\frac{W}{L}\right)_p(V_{DD}-|V_{tp}|)\right]$$



Inverter VTC and Noise Margins

$$V_M = \frac{r(V_{DD}-|V_{tp}|)+V_{tn}}{1+r} \quad \text{where} \quad r = \sqrt{\frac{k_p'(W/L)_p}{k_n'(W/L)_n}}$$

For matched devices, that is, $\mu_n\left(\frac{W}{L}\right)_n = \mu_p\left(\frac{W}{L}\right)_p$, and $V_{tn} = -V_{tp} = V_t$

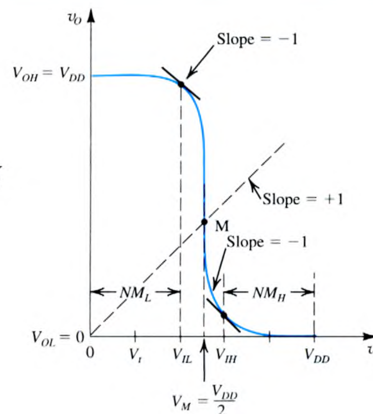
$r = 1$

$V_M = V_{DD}/2$

$V_{IL} = \frac{1}{8}(3V_{DD}+2V_t)$

$V_{IH} = \frac{1}{8}(5V_{DD}-2V_t)$

$NM_H = NM_L = \frac{1}{8}(3V_{DD}+2V_t)$



Computer Simulation Problems

SIM Problems identified by the LTSpice icon are intended to demonstrate the value of using SPICE simulation to verify hand analysis and design, and to investigate important issues such as gate noise margins and propagation delays. Instructions to assist in setting up simulations for all the indicated problems can be found in the corresponding files on the website. Note that if a particular parameter value is not specified in the problem statement, you are to make a reasonable assumption.

Section 16.1: CMOS Logic-Gate Circuits

16.1 Consider MOS transistors fabricated in a 65-nm process for which $\mu_n C_{ox} = 540\ \mu\text{A/V}^2$, $\mu_p C_{ox} = 100\ \mu\text{A/V}^2$, $V_{tn} = -V_{tp} = 0.35$ V, and $V_{DD} = 1$ V.

(a) Find R_{on} of an NMOS transistor with $W/L = 1.5$.
(b) Find R_{on} of a PMOS transistor with $W/L = 1.5$.
(c) If R_{on} of the PMOS device is to be equal to that of the NMOS device in (a), what must $(W/L)_p$ be?

16.2 The CMOS inverter of Fig. 16.2(b) is implemented in a 0.13-μm process for which $\mu_n C_{ox} = 500\ \mu\text{A/V}^2$, $\mu_p C_{ox} = 125\ \mu\text{A/V}^2$, $V_{tn} = -V_{tp} = 0.4$ V, and $V_{DD} = 1.2$ V. The NMOS transistor has $(W/L)_n = 1.5$.

(a) What must $(W/L)_p$ be if Q_N and Q_P are to have equal R_{on} resistances?
(b) Find the value of R_{on}.

16.3 Give the CMOS circuit that realizes a three-input NOR gate.

16.4 Give the CMOS circuit for a three-input NAND gate.

16.5 Find the PUN that corresponds to the PDN shown in Fig. P16.5, and hence the complete CMOS logic circuit. What is the Boolean function realized?

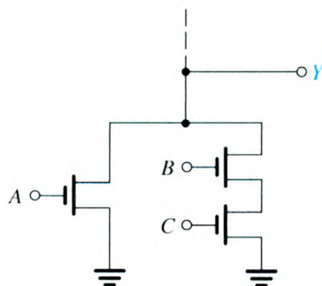


Figure P16.5

16.6 Find the PUN that corresponds to the PDN shown in Fig. P16.6, and hence the complete CMOS logic circuit. What is the Boolean function realized?

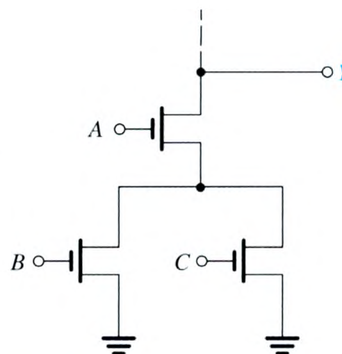


Figure P16.6

16.7 Find the PDN that corresponds to the PUN shown in Fig. P16.7, and hence the complete CMOS logic circuit. What is the Boolean function realized?

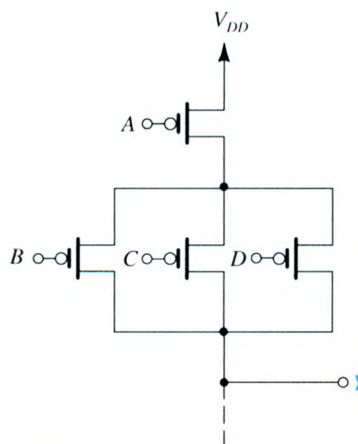


Figure P16.7

16.8 Give the CMOS realization for the Boolean function

$$Y = \overline{AB + CDE}$$

16.9 Find the PDN that is the dual of the PUN in Fig. 16.10(a) and hence give a CMOS realization of the exclusive-OR (XOR) function.

D 16.10 Provide a CMOS logic gate circuit that realizes the function

$$Y = \overline{A}BC + A\overline{B}C + AB\overline{C}$$

How many transistors are required? Explore the possibility of reducing the number of the transistors required.

D 16.11 Sketch a CMOS logic circuit that realizes the function $Y = AB + \overline{A}\,\overline{B}$. This is called the **equivalence** or **coincidence function**.

D 16.12 Sketch a CMOS logic circuit that realizes the function $Y = ABC + \overline{A}\,\overline{B}\,\overline{C}$.

D*16.13 It is required to design a CMOS logic circuit that realizes a three-input, even-parity checker. Specifically, the output Y is to be low when an even number (0 or 2) of the inputs A, B, and C are high.

(a) Give the Boolean function $\overline{Y}$.
(b) Sketch a PDN directly from the expression for $\overline{Y}$. Note that it requires 12 transistors in addition to those in the inverters.
(c) From inspection of the PDN circuit, reduce the number of transistors to 10 (not counting those in the inverters).
(d) Find the PUN as a dual of the PDN in (c), and hence the complete realization.

D *16.14 Give a CMOS logic circuit that realizes the function of a three-input, odd-parity checker. Specifically, the output is to be low when an odd number (1 or 3) of the inputs are high. Attempt a design with 10 transistors (not counting those in the inverters) in each of the PUN and the PDN.

D *16.15 Design a CMOS full-adder circuit with inputs A, B, and C, and two outputs S and C_0 such that S is 1 if one or three inputs are 1, and C_0 is 1 if two or more inputs are 1.

Section 16.2: Digital Logic Inverters

16.16 A particular logic inverter is specified to have $V_{IL} = 0.5$ V, $V_{IH} = 0.7$ V, $V_{OL} = 0.1$ V, and $V_{OH} = 1.2$ V. Find the high and low noise margins, NM_H and NM_L.

16.17 The voltage-transfer characteristic of a particular logic inverter is modeled by three straight-line segments in the manner shown in Fig. 16.13. If $V_{IL} = 1.1$ V, $V_{IH} = 1.2$ V, $V_{OL} = 0.3$ V, and $V_{OH} = 1.5$ V, find:

(a) the noise margins
(b) the value of V_M
(c) the voltage gain in the transition region

16.18 For a particular inverter design using a power supply V_{DD}, $V_{OL} = 0.1V_{DD}$, $V_{OH} = 0.8V_{DD}$, $V_{IL} = 0.25V_{DD}$, and $V_{IH} = 0.6\,V_{DD}$. What are the noise margins? What is the width of the transition region? For a minimum noise margin of 0.25 V, what value of V_{DD} is required?

***16.19** Consider an inverter implemented as in Fig. 16.17. Let $V_{DD} = 1.8$ V, $R = 2$ kΩ, $R_{on} = 100$ Ω, $V_{IL} = 0.6$ V, and $V_{IH} = 0.8$ V.

(a) Find V_{OL}, V_{OH}, NM_H, and NM_L.
(b) The inverter is driving N identical inverters. Each of these load inverters, or **fan-out** inverters as they are usually called, is specified to require an input current of 0.1 mA when the input voltage (of the fan-out inverter) is high and zero current when the input voltage is low. Noting that the input currents of the fan-out inverters will have to be supplied through R of the driving inverter, find the resulting value of V_{OH} and of NM_H as a function of the number of fan-out inverters N. Hence find the maximum value N can have while the inverter is still providing an NM_H value that is greater than or equal to its NM_L.
(c) Find the power dissipation in the inverter in the two cases: (i) the output is low, and (ii) the output is high and driving the maximum fan-out found in (b).

16.20 For an inverter employing a 1.8-V supply, suggest an ideal set of values for V_M, V_{IL}, V_{IH}, V_{OL}, V_{OH}, NM_L, NM_H. Also, sketch the VTC. What value of voltage gain in the transition region does your ideal specification imply?

***16.21** For a particular inverter, the basic technology used provides an inherent limit to the small-signal, low-frequency voltage gain of 50 V/V. If, with a 1-V supply, the values of V_{OL} and V_{OH} are ideal, but $V_M = 0.4V_{DD}$, what are the best possible values of V_{IL} and V_{IH} that can be expected? What are the best possible noise margins you could expect? Find the large-signal voltage gain, where the gain is defined by $(V_{OH} - V_{OL})/(V_{IL} - V_{IH})$. (*Hint:* Use straight-line approximations for the VTC.)

***16.22** A logic-circuit type intended for use in a digital-signal-processing application in a newly developed

hearing aid can operate down to single-cell supply voltages of 0.9 V. If for its inverter, the output signals swing between 0 and V_{DD}, the "gain-of-one" points are separated by less than $\frac{1}{3}V_{DD}$, and the noise margins are within 30% of one another, what ranges of values of V_{IL}, V_{IH}, V_{OL}, V_{OH}, NM_L, and NM_H can you expect for the lowest possible battery supply?

D 16.23 Design the inverter circuit in Fig. 16.12(a) to provide $V_{OH} = 1.0$ V, $V_{OL} = 50$ mV, and so that the current drawn from the supply in the low-output state is 30 μA. The transistor has $V_t = 0.35$ V, $\mu_n C_{ox} = 540\ \mu\text{A/V}^2$, and $\lambda = 0$. Specify the required values of V_{DD}, R_D, and W/L. How much power is drawn from the supply when the output is high? When the output is low?

D 16.24 Refer to the analysis of the resistive-load MOS inverter in Example 16.2 and utilize the expressions derived there for the various inverter parameters. Design the circuit to satisfy the following requirements: $V_{OH} = 1.2$ V, $V_{OL} = 50$ mV, and the power dissipation in the low-output state = 50 μW. The transistor available has $V_t = 0.4$ V, $\mu_n C_{ox} = 500\ \mu\text{A/V}^2$, and $\lambda = 0$. Specify the required values of V_{DD}, R_D, and W/L. What are the values obtained for V_{IL}, V_M, V_{IH}, NM_L, and NM_H?

D *16.25 Refer to the analysis of the resistive-load MOS inverter in Example 16.2 and utilize the expressions derived there for the various inverter parameters. For a technology for which $V_t = 0.3V_{DD}$, it is required to design the inverter to obtain $V_M = V_{DD}/2$. In terms of V_{DD}, what is the required value of the design parameter V_x? What values are obtained for V_{OH}, V_{OL}, V_{IL}, V_{IH}, NM_H, and NM_L, in terms of V_{DD}? Give numerical values for the case $V_{DD} = 1.0$V. Now, express the power dissipated in the inverter in its low-output state in terms of the transistor's W/L ratio. Let $k_n' = 540\ \mu\text{A/V}^2$. If the power dissipation is to be limited to approximately 100 μW, what W/L ratio is needed and what value of R_D is used?

***16.26** Consider a pseudo-NMOS inverter as shown in Fig. 16.21 and fabricated in a 65-nm CMOS technology for which $V_{DD} = 1.0$ V, $|V_t| = 0.35$ V, $k_n/k_p = 5.4$, and $k_n = 540\ \mu\text{A/V}^2$.

(a) Find V_{OH}.
(b) Derive an equation for V_{OL} that is a function of V_{DD}, V_t, and k_n/k_p.
(c) Evaluate the equation using the process parameters in the problem statement.

****16.27** Derive an expression for V_M of the pseudo-NMOS inverter shown in Fig. 16.21. You may assume that $V_t = V_{tn} = |V_{tp}|$ and $r \equiv k_n/k_p$.

Section 16.3: The CMOS Inverter

16.28 Consider a CMOS inverter fabricated in a 65-nm CMOS process for which $V_{DD} = 1$ V, $V_{tn} = -V_{tp} = 0.35$ V, and $\mu_n C_{ox} = 5.4\mu_p C_{ox} = 540\ \mu\text{A/V}^2$. In addition, Q_N and Q_P have $L = 65$ nm and $(W/L)_n = 1.5$.

(a) Find W_p that results in $V_M = V_{DD}/2$. What is the silicon area utilized by the inverter in this case?
(b) For the matched case in (a), find the values of V_{OH}, V_{OL}, V_{IH}, V_{IL}, NM_L, and NM_H.
(c) For the matched case in (a), find the output resistance of the inverter in each of its two states.

16.29 Consider a CMOS inverter fabricated in a 0.25-μm CMOS process for which $V_{DD} = 2.5$ V, $V_{tn} = -V_{tp} = 0.5$ V, and $\mu_n C_{ox} = 3.5\ \mu_p C_{ox} = 115\ \mu\text{A/V}^2$. In addition, Q_N and Q_P have $L = 0.25\ \mu$m and $(W/L)_n = 1.5$. Investigate the variation of V_M with the ratio W_p/W_n. Specifically, calculate V_M for (a) $W_p = 3.5W_n$ (the matched case), (b) $W_p = W_n$ (the minimum-size case); and (c) $W_p = 2W_n$ (a compromise case). For cases (b) and (c), estimate the approximate reduction in NM_L and silicon area relative to the matched case (a).

***16.30** For a technology in which $V_{tn} = 0.3V_{DD}$, show that the maximum current that the inverter can sink while its low-output level does not exceed 0.1 V_{DD} is $0.065\ k_n'(W/L)_n V_{DD}^2$. For $V_{DD} = 1.2$V, $k_n' = 500\ \mu\text{A/V}^2$, find $(W/L)_n$ that permits this maximum current to be 0.1 mA.

16.31 A CMOS inverter for which $k_n = 5k_p = 250\ \mu\text{A/V}^2$ and $V_t = 0.4$ V is connected as shown in Fig. P16.31 to a sinusoidal signal source having a Thévenin equivalent voltage of 0.1-V peak amplitude and resistance of 100 kΩ. What signal voltage appears at node A with $v_I = +1.5$ V? With $v_I = -1.5$ V?

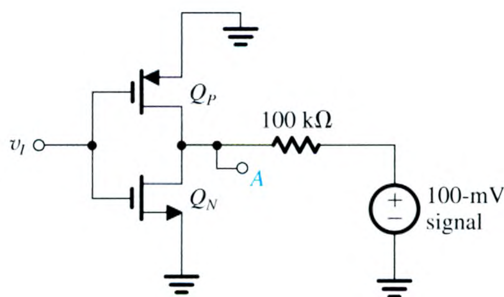


Figure P16.31

D 16.32 There are situations in which Q_N and Q_P of the CMOS inverter are deliberately mismatched to realize a

certain desired value for V_M. Show that the value required of the parameter r of Eq. (16.40) is given by

$$r = \frac{V_M - V_{tn}}{V_{DD} - |V_{tp}| - V_M}$$

For a 65-nm process characterized by $V_{tn} = -V_{tp} = 0.35\,\text{V}$, $V_{DD} = 1.0\,\text{V}$, and $\mu_n = 5.4\mu_p$, find the ratio W_p/W_n required to obtain $V_M = 0.6V_{DD}$.

16.33 Consider the CMOS inverter of Fig. 16.22 with Q_N and Q_P matched and with the input v_I rising slowly from 0 to V_{DD}. At what value of v_I does the current flowing through Q_N and Q_P reach its peak? Give an expression for the peak current, neglecting λ_n and λ_p. For $k_n' = 540\ \mu\text{A/V}^2$, $(W/L)_n = 1.5$, $V_{DD} = 1.0$ V, and $V_{tn} = 0.35\,\text{V}$, find the value of the peak current.

16.34 Repeat Example 16.3 for a CMOS inverter fabricated in a 0.13-μm process for which $V_{DD} = 1.3$ V, $V_{tn} = |V_{tp}| = 0.4$ V, $\mu_n = 4\mu_p$, and $\mu_n C_{ox} = 500\ \mu\text{A/V}^2$. In addition, Q_N and Q_P have $L = 0.13$ μm and $(W/L)_n = 1.5$. For part (a) use $V_M = V_{DD}/2 = 0.65$ V.

16.35 Redo Exercise 16.5 using the following parameters from a 28-nm CMOS technology: $V_{DD} = 0.9$ V, $V_{tn} = -V_{tp} = 0.3$ V, $\mu_n/\mu_p = 1.5$, and $\mu_n C_{ox} = 750\ \mu\text{A/V}^2$. Q_N and Q_P have $L = 28$ nm and $(W/L)_n = 2.0$.

CHAPTER 17

Digital Design: Power, Speed, and Area

IN THIS CHAPTER YOU WILL LEARN

- How to analyze the dynamic performance of a CMOS inverter.
- How to select sizes for the transistors in a CMOS logic circuit so as to meet various performance requirements.
- The sources of power consumption in logic circuits, with emphasis on CMOS, and the trade-off between power dissipation and speed of operation.
- The implications of technology scaling (Moore's law) over more than 50 years, and some of the current challenges and opportunities in the design of nanoscale digital circuits.

Introduction

In Chapter 16, we studied the basic principles underlying the construction and static operation of CMOS logic circuits. Having done this, we can now turn our attention to the three most important design considerations for digital circuits: speed, power, and area. High-speed circuits are often needed to meet consumer demands, low-power operation is often required to meet thermal or battery constraints, and circuit area has a very significant impact on cost. Therefore, optimizing each of these metrics is extremely important. However, as we will see, optimizing any one of these metrics often comes at a cost to the other two—an unfortunate yet common circumstance in any field of engineering! Consider speed, for instance. We will see in this chapter that increasing current by making a transistor larger can improve speed; however, the larger transistor requires greater area and, owing to the larger current and parasitic capacitances, more power.

This chapter provides an overview of design and optimization approaches for combinational CMOS logic circuits, with a focus on these three metrics. We will explore several interrelated aspects, with a focus on analyzing and optimizing speed of operation (Section 17.1), selecting appropriate transistor sizes (Section 17.2), and identifying and characterizing sources of power consumption (Section 17.3). This lays the groundwork for the study of sequential and memory circuits in Chapter 18.

Section 17.4 looks at the implications of Moore's law on digital integrated circuit design. Specifically, over the last 50 years or so, MOSFET dimensions have been reduced by a factor of 2 every 5 years. This scaling has been accompanied by reductions in V_{DD} and V_t. We will consider both the opportunities and the challenges that come with scaling, especially in terms of speed, area, and power. Finally, we will offer some insights to assist designers in adapting their knowledge to emerging technologies and logic circuit families not covered in this textbook.

17.1 Dynamic Operation of the CMOS Inverter

The speed of operation of a digital system (e.g., a computer) is determined by the propagation delay of the logic gates used to construct the system. Since the inverter is the basic logic gate of any digital IC technology, the propagation delay of the inverter is a fundamental parameter in characterizing the speed of a given technology. We begin our study of the dynamic operation of CMOS in Section 17.1.1 by considering the propagation delay of a general inverter circuit. There, we introduce key definitions and analysis methods that are applied in the CMOS case in Sections 17.1.2 and 17.1.3.

17.1.1 Propagation Delay

The propagation delay is the time the inverter takes to respond to a change at its input. To be specific, let us consider an inverter fed with the ideal pulse shown in Fig. 17.1(a). The resulting output signal of the inverter is shown in Fig. 17.1(b). Notice that:

1. The output signal is no longer an ideal pulse. Rather, it has rounded edges; that is, the pulse takes some time to fall to its low value and to rise to its high value. In

technical terms, we say that the pulse has finite fall and rise times. We will provide a precise definition of these shortly.

2. There is a time delay between each edge of the input pulse and the corresponding change in the output of the inverter. If we define the "switching point" of the output as the time at which the output pulse passes through the half-point of its excursion, then we can define the propagation delays of the inverter as indicated in Fig. 17.1(b). Note that there are two propagation delays, which are not necessarily equal: the propagation delay for the output going from high to low, t_{PHL}, and the propagation delay for the output going from low to high, t_{PLH}. The inverter propagation delay t_P is defined as the average of the two,

$$t_P \equiv \frac{1}{2}(t_{PLH} + t_{PHL}) \tag{17.1}$$

At this point, you are likely wondering about the cause of the finite propagation time of the inverter. It is simply a result of the time needed to charge and discharge the various capacitances in the circuit. These include the MOSFET capacitances, the wiring capacitance, and the input capacitances of all the logic gates driven by the inverter. We will have a lot more to say about these capacitances and about the determination of t_P shortly. For the time being, however, be aware that

1. A fundamental relationship in analyzing the dynamic operation of a circuit is

$$I\Delta t = \Delta Q = C\Delta V \tag{17.2}$$

That is, a current I flowing through a capacitance C for an interval Δt deposits a charge ΔQ on the capacitor, which causes the capacitor voltage to increase by ΔV.

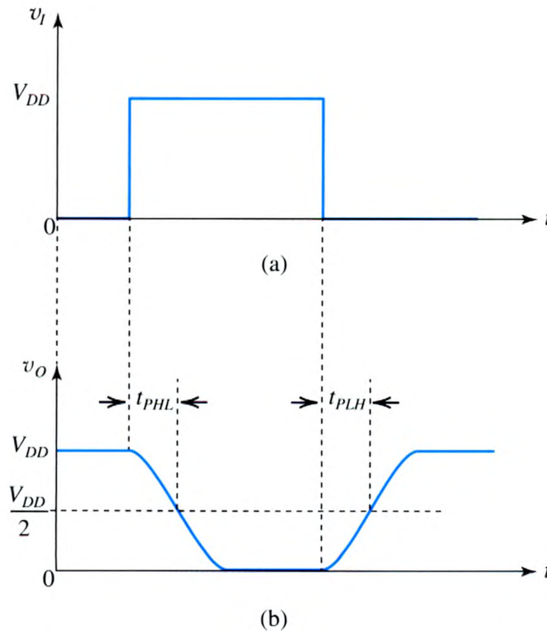


Figure 17.1 An inverter fed with the ideal pulse in **(a)** provides at its output the pulse in **(b)**. Two delay times are defined as indicated.

2. A thorough familiarity with the time response of single-time-constant (STC) circuits is of great help in the analysis of the dynamic operation of digital circuits. A review of this subject is presented in Appendix E. For our purposes here, we remind you of the key equation in determining the response to a step function. Consider a step-function input applied to an STC circuit, either low-pass or high-pass, and let the circuit have a time constant τ. The output at any time t is given by

$$y(t) = Y_{\infty} - \left(Y_{\infty} - Y_{0+}\right)e^{-t/\tau} \quad (17.3)$$

where Y_{∞} is the final value, that is, the value toward which the response is heading, and Y_{0+} is the value of the response immediately after $t = 0$. This equation states that the output at any time t is equal to the difference between the final value Y_{∞} and a gap whose initial value is $Y_{\infty} - Y_{0+}$ and that is shrinking exponentially.

Example 17.1 Calculating the Propagation Delay of a Simple Inverter

For the inverter in Fig. 17.2(a), consider the case where a capacitor C is connected between the output node and ground. If at $t = 0$, v_I goes low, and assuming that the transistor turns off instantaneously, find the time for v_O to reach $\frac{1}{2}(V_{OH} + V_{OL})$. This is the low-to-high propagation time, t_{PLH}. Calculate the value of t_{PLH} for the case $R = 25\,\text{k}\Omega$ and $C = 8$ fF.

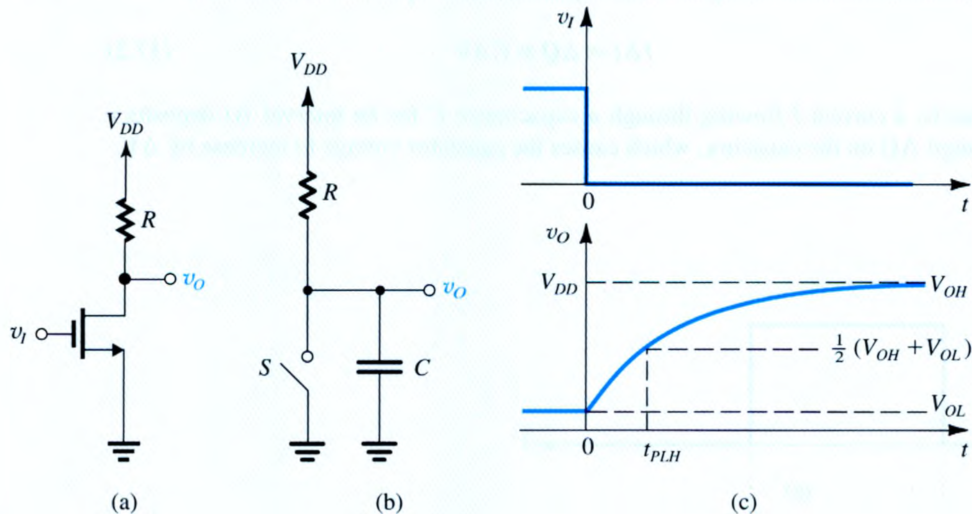


Figure 17.2 Example 17.1: **(a)** Inverter schematic circuit. **(b)** The inverter circuit after v_I goes low and the switch opens (i.e., for $t > 0$). **(c)** Waveforms of v_I and v_O. Observe that the switch is assumed to operate instantaneously. v_O rises exponentially, starting at V_{OL} and heading toward V_{OH}.

Solution

While the transistor is on, $v_O = V_{OL}$. When the transistor turns off at $t = 0$, the circuit takes the form shown in Fig. 17.2(b). Since the voltage across the capacitor cannot change instantaneously, at $t = 0+$

the output will still be V_{OL}. Then the capacitor charges through R, and v_O rises exponentially toward V_{DD}. The output waveform will be as shown in Fig. 17.2(c), and its equation can be obtained by substituting in Eq. (17.3): $v_O(\infty) = V_{OH} = V_{DD}$ and $v_O(0+) = V_{OL}$. Thus,

$$v_O(t) = V_{OH} - (V_{OH} - V_{OL})e^{-t/\tau}$$

where $\tau = CR$. To find t_{PLH}, we substitute

$$v_O(t_{PLH}) = \frac{1}{2}(V_{OH} + V_{OL})$$

Thus,

$$\frac{1}{2}(V_{OH} + V_{OL}) = V_{OH} - (V_{OH} - V_{OL})e^{-t_{PLH}/\tau}$$

which results in

$$t_{PLH} = \tau \ln 2 = 0.69\tau$$

Note that this expression is independent of the values of V_{OL} and V_{OH}. For the numerical values given,

$$\begin{aligned} t_{PLH} &= 0.69RC \\ &= 0.69 \times 25 \times 10^3 \times 8 \times 10^{-15} \\ &= 138 \text{ ps} \end{aligned}$$

EXERCISES

17.1 A capacitor C whose initial voltage is 0 is charged to a voltage V_{DD} by a constant-current source I. Find the time t_{PLH} at which the capacitor voltage reaches $(V_{DD}/2)$. What value of I is required to obtain a 10-ps propagation delay with $C = 10$ fF and $V_{DD} = 1.2$ V?
Ans. $t_{PLH} = CV_{DD}/2I$; 0.6 mA

17.2 For the inverter of Fig. 16.18(a), let the on-resistance of P_U be 20 kΩ and that of $P_D = 10$ kΩ. If the capacitance C between the output and ground is 10 fF, find t_{PLH}, t_{PHL}, and t_P.
Ans. 138 ps; 69 ps; 104 ps

We conclude this section by showing in Fig. 17.3 the formal definition of the propagation delay of an inverter. As shown, an input pulse with finite (nonzero) **rise and fall times** is applied. The inverted pulse at the output exhibits finite rise and fall times (labeled t_{TLH} and t_{THL}, where the subscript T denotes transition, LH denotes low to high, and HL denotes

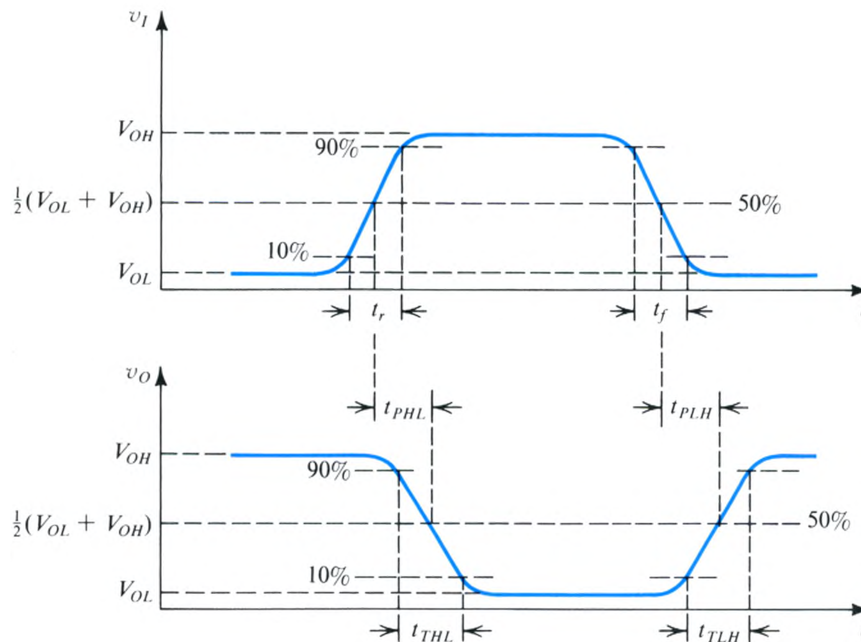


Figure 17.3 Definitions of propagation delays and transition times of the logic inverter.

high to low). There is also a delay time between the input and output waveforms. As given in Eq. (17.1), the usual way to specify the propagation delay is to take the average of the high-to-low propagation delay, t_{PHL}, and the low-to-high propagation delay, t_{PLH}. As indicated, these delays are measured between the 50% points of the input and output waveforms. Also note that the **transition times** are specified using the 10% and 90% points of the output excursion ($V_{OH} - V_{OL}$).

EXERCISE

17.3 A capacitor $C = 100$ fF is discharged from a voltage V_{DD} to zero through a resistance $R = 2$ kΩ. Find the fall time t_f of the capacitor voltage.
Ans. $t_f \simeq 2.2CR = 0.44$ ns

17.1.2 Determining the Propagation Delay of the CMOS Inverter

Our strategy for determining the propagation delay of the CMOS inverter consists of two steps:

1. Replace all the capacitances in the circuit; that is, the various capacitances associated with Q_N and Q_P, the capacitance of the wire that connects the output of the inverter

to other circuits, and the input capacitance of the logic gates the inverter drives, by a single equivalent capacitance C connected between the output node of the inverter and ground.

2. Analyze the resulting capacitively loaded inverter to determine its t_{PLH} and t_{PHL}, and hence t_P.

We shall study these two separable steps in reverse order. Thus, in this section we show how the propagation delay can be determined. Then, in Section 17.1.3, we show how to calculate the value of C.

Figure 17.4(a) shows a CMOS inverter with a capacitance C connected between its output node and ground. To determine the propagation delays t_{PHL} and t_{PLH}, we apply to the input an ideal pulse, that is, one with zero rise and fall times, as shown in Fig. 17.4(b). Since the circuit has a symmetric structure, the analyses to determine the two propagation delays will be similar. Therefore, we will derive t_{PHL} in detail and extrapolate the result to determine t_{PLH}.

Just prior to the leading edge of the input pulse (i.e., at $t = 0-$), the output voltage is equal to V_{DD} and capacitor C is charged to this voltage. At $t = 0$, v_I rises to V_{DD}, causing Q_P to turn off and Q_N to turn on. From then on, the circuit is equivalent to that shown in Fig. 17.4(c), with the initial value of $v_O = V_{DD}$. Thus, at $t = 0+$, Q_N will operate in the saturation region and will supply a relatively large current to begin the process of discharging C. Figure 17.4(d) shows the trajectory of the operating point of Q_N as C is discharged. Here we are interested in the interval t_{PHL} during which v_O reduces from V_{DD} to $V_{DD}/2$. Correspondingly, the operating point of Q_N moves from E to M. For a portion of this time, corresponding to the segment EF of the trajectory, Q_N operates in saturation. Then at F, $v_O = V_{DD} - V_t$, and Q_N enters the triode region.

A simple approach for determining t_{PHL} consists of first calculating the average value of the current supplied by Q_N over the segment EM. Then, we use this average value of the discharge current to approximate t_{PHL} by means of the charge balance equation

$$I_{av}\, t_{PHL} = C\left[V_{DD} - (V_{DD}/2)\right]$$

resulting in

$$t_{PHL} = \frac{CV_{DD}}{2I_{av}} \tag{17.4}$$

The value of I_{av} can be found as follows:

$$I_{av} = \frac{1}{2}\left[i_{DN}(\text{E}) + i_{DN}(\text{M})\right] \tag{17.5}$$

where

$$i_{DN}(\text{E}) = \frac{1}{2}k_n'\left(\frac{W}{L}\right)_n (V_{DD} - V_{tn})^2 \tag{17.6}$$

and

$$i_{DN}(\text{M}) = k_n'\left(\frac{W}{L}\right)_n \left[(V_{DD} - V_{tn})\left(\frac{V_{DD}}{2}\right) - \frac{1}{2}\left(\frac{V_{DD}}{2}\right)^2\right] \tag{17.7}$$

Note that we have assumed $\lambda_n = 0$. Combining Eqs. (17.4) to (17.7) provides

$$t_{PHL} = \frac{\alpha_n C}{k_n'(W/L)_n V_{DD}} \tag{17.8}$$

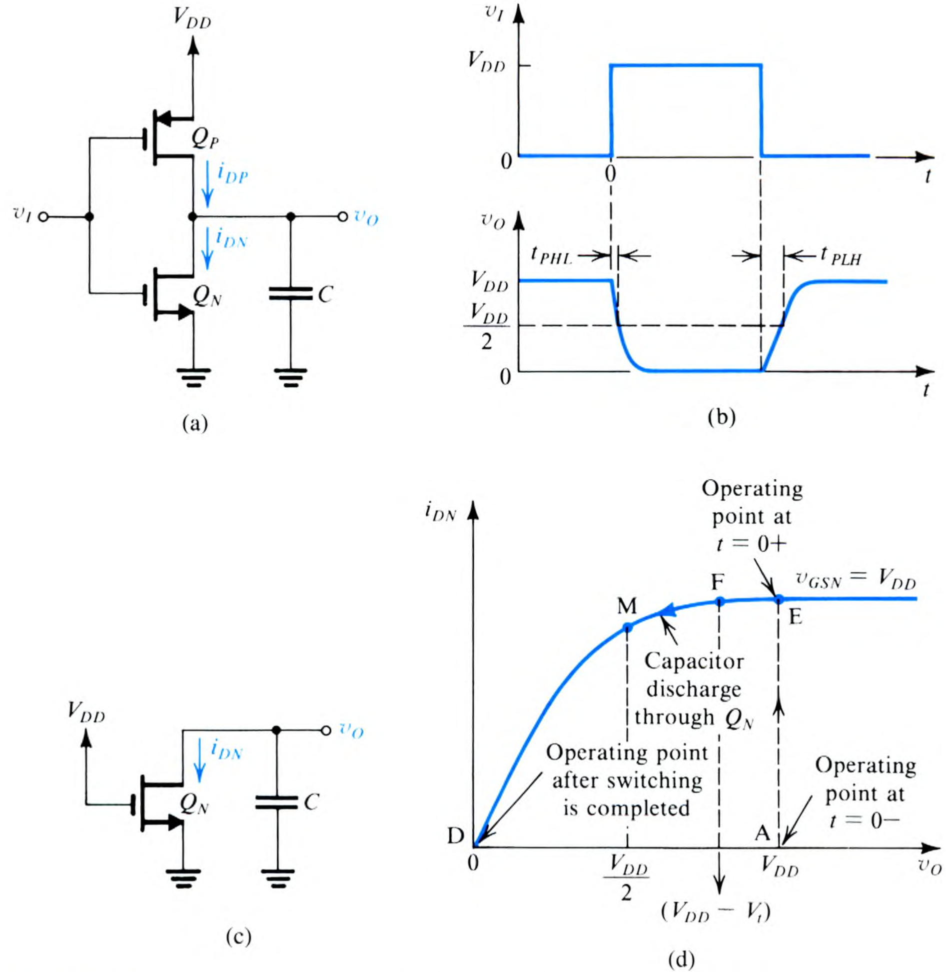


Figure 17.4 Dynamic operation of a capacitively loaded CMOS inverter: **(a)** circuit; **(b)** input and output waveforms; **(c)** equivalent circuit during the capacitor discharge; **(d)** trajectory of the operating point as the input goes high and C discharges through Q_N.

where α_n is a factor determined by the relative values of V_t and V_{DD};

$$\alpha_n = 2\Bigg/\left[\frac{7}{4} - \frac{3V_{tn}}{V_{DD}} + \left(\frac{V_{tn}}{V_{DD}}\right)^2\right] \tag{17.9}$$

The value of α_n typically falls in the range of 1 to 3.

An expression for the low-to-high inverter delay, t_{PLH}, can be written by analogy to the t_{PHL} expression in Eq. (17.8),

$$t_{PLH} = \frac{\alpha_p C}{k'_p (W/L)_p V_{DD}} \tag{17.10}$$

where

$$\alpha_p = 2 \Bigg/ \left[\frac{7}{4} - \frac{3|V_{tp}|}{V_{DD}} + \left| \frac{V_{tp}}{V_{DD}} \right|^2 \right] \tag{17.11}$$

Finally, the propagation delay t_P can be found as the average of t_{PHL} and t_{PLH}, as in Eq. (17.1). Examination of the formulas in Eqs. (17.8) to (17.11) enables us to make a number of useful observations:

1. As expected, the two components of t_P can be equalized by selecting the (W/L) ratios to equalize k_n and k_p, that is, by matching Q_N and Q_P. This assumes that $\alpha_n = \alpha_p$, which is obtained when $V_{tn} = -V_{tp}$.
2. Since t_P is proportional to C, the designer should strive to reduce C. This is achieved by using the minimum possible channel length and by minimizing wiring and other parasitic capacitances. Careful layout of the chip can result in significant reduction in such capacitances.
3. Using a process technology with larger transconductance parameter k' can result in shorter propagation delays. Keep in mind, however, that for such processes C_{ox} is increased, and thus the value of C increases at the same time (more on this in Section 17.4).
4. Using larger W/L ratios can result in a reduction in t_P. Care should be taken here, however, since increasing the size of the devices increases the value of C, so that the expected reduction in t_P might not be possible. Reducing t_P by increasing W/L, however, is an effective strategy when C is dominated by components not directly related to the size of the driving device (such as wiring or fan-out devices).
5. A larger supply voltage V_{DD} results in a lower t_P. However, V_{DD} is determined by the process technology and thus is often not under the control of the designer. Furthermore, modern process technologies in which device sizes are reduced require lower V_{DD} (see Appendix K). A motivating factor for lowering V_{DD} is the need to keep the dynamic power dissipation at acceptable levels, especially in very-high-density chips. We will have more to say on this point in Section 17.3.

These observations clearly illustrate the conflicting requirements and the trade-offs available in the design of a CMOS digital integrated circuit.

An Alternative Approach The formulas derived above for t_{PHL} and t_{PLH} underestimate the delay values for inverters implemented in deep-submicron technologies. This arises because of the velocity-saturation effect, which we shall discuss briefly in Section 17.4. Velocity saturation results in lower MOSFET currents than are predicted in the saturation region, and hence in increased delay times. To deal with this problem, we present a very simple alternative approach to estimating the inverter propagation delay.

Figure 17.5 illustrates the alternative approach. During the discharge delay t_{PHL}, Q_N is replaced by an equivalent resistance R_N. Similarly, during the charging delay t_{PLH}, Q_P is replaced by an equivalent resistance R_P. It is easy to show that

$$t_{PHL} = 0.69 R_N C \tag{17.12}$$

and

$$t_{PLH} = 0.69 R_P C \tag{17.13}$$

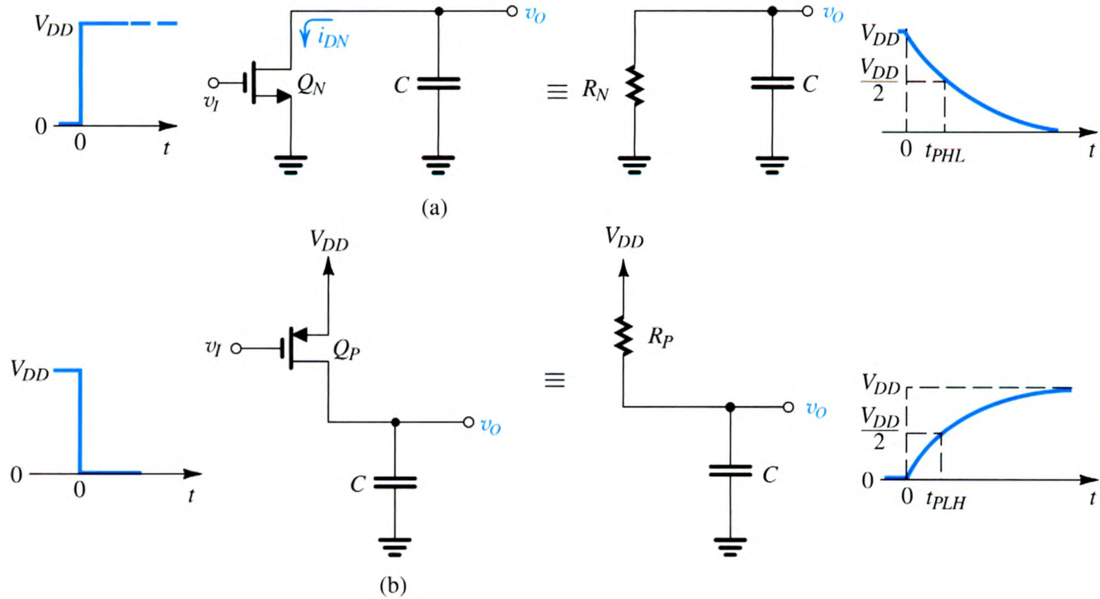


Figure 17.5 Equivalent circuits for determining the propagation delays **(a)** t_{PHL} and **(b)** t_{PLH} of the inverter.

Values for R_N and R_P can be estimated by simulation or found empirically. Often, when using a new technology, a designer (who is often part of a larger design team) simulates the circuits of Fig. 17.5 using "unit-sized" NMOS and PMOS transistors (i.e., transistors with known dimensions W_{eff} and L_{eff}) and a large external load capacitance C (chosen so that it dominates over the transistor's internal capacitances). Based on these simulations, values of t_{PHL} and t_{PLH} can be obtained, and hence, effective values of R_N and R_P for that particular transistor size, which we denote as as $R_{eff,N}$ and $R_{eff,P}$, can be determined based on Eqs. (17.12) and (17.13). Following this, values of R_N and R_P for transistors with dimensions W_n/L_n and W_p/L_p, respectively, can be estimated as:

$$R_N = \frac{R_{eff,N}(W_{eff}/L_{eff})}{(W_n/L_n)} \tag{17.14}$$

$$R_P = \frac{R_{eff,P}(W_{eff}/L_{eff})}{(W_p/L_p)} \tag{17.15}$$

Depending on the technology and transistor dimensions, values of R_N and R_P typically range from tens of Ω to several kΩ. For CMOS technologies including 0.25 μm, 0.18 μm, and 0.13 μm, it has been found that values of $R_{eff,N}$ = 12.5 kΩ and $R_{eff,P}$ = 30 kΩ are quite accurate for $W_{eff}/L_{eff} = 1$ (see Hodges et al., 2004).

Finally, note that the delay expressions in Eqs. (17.12) and (17.13) are obtained by assuming that the inverter is driven by a step-input voltage. In the more practical case of a ramp-input voltage, it has been shown that the 0.69 factor approaches unity, thus

$$t_{PHL} \simeq R_N C \tag{17.16}$$

and

$$t_{PLH} \simeq R_P C \tag{17.17}$$

Example 17.2 Determining the Propagation Delay of the CMOS Inverter

For a 65-nm process characterized by $V_{DD} = 1.0$ V, $V_{tn} = -V_{tp} = 0.35$ V, $k_n' = 5.4k_p' = 540\ \mu A/V^2$, find t_{PLH}, t_{PHL}, and t_P for an inverter for which $(W/L)_n = 1.5$ and $(W/L)_p = 3$, and for $C = 5$ fF. In this technology, transistors with $(W/L) = 1.5$ are found to have $R_{eff,N} = 8$ kΩ and $R_{eff,P} = 24$ kΩ. Use both the approach based on average currents and the one based on equivalent resistances, and compare your results. If to save on power dissipation the inverter is operated at $V_{DD} = 0.8$ V, by what factor does t_P change?

Solution

(a) Using the average current approach, we determine, from Eq. (17.9),

$$\alpha_n = \frac{2}{\frac{7}{4} - \frac{3\times 0.35}{1.0} + \left(\frac{0.35}{1.0}\right)^2} = 2.43$$

and using Eq. (17.8),

$$t_{PHL} = \frac{2.43 \times 5 \times 10^{-15}}{540 \times 10^{-6} \times 1.5 \times 1.0} = 15.0 \text{ ps}$$

Since $|V_{tp}| = V_{tn}$,

$$\alpha_p = \alpha_n = 2.43$$

and we can determine t_{PLH} from Eq. (17.10) as

$$t_{PLH} = \frac{2.43 \times 5 \times 10^{-15}}{(540/5.4) \times 10^{-6} \times 3 \times 1.0} = 40.5 \text{ ps}$$

The propagation delay can now be found as

$$t_P = \frac{1}{2}(t_{PHL} + t_{PLH}) = \frac{1}{2}(15.0 + 40.5) = 27.8 \text{ ps}$$

(b) Using the equivalent-resistance approach, we first find R_N from Eq. (17.14) as

$$R_N = \frac{8 \times 1.5}{1.5} = 8 \text{ k}\Omega$$

Example 17.2 *continued*

and then use Eq. (17.12) to determine t_{PHL},

$$t_{PHL} = 0.69 \times 8 \times 10^3 \times 5 \times 10^{-15} = 27.6 \text{ ps}$$

Similarly, we use Eq. (17.15) to determine R_P,

$$R_P = \frac{24 \times 1.5}{3} = 12 \text{ k}\Omega$$

and Eq. (17.13) to determine t_{PLH},

$$t_{PLH} = 0.69 \times 12 \times 10^3 \times 5 \times 10^{-15} = 41.4 \text{ ps}$$

Thus, while the value obtained for t_{PHL} is higher than that found using average currents, the value for t_{PLH} is about the same. Finally, t_P can be found as

$$t_P = \frac{1}{2}(27.6 + 41.4) = 34.5 \text{ ps}$$

which is a little higher than the value we found using average currents.

To find the change in propagation delays obtained when the inverter is operated at $V_{DD} = 0.8$ V, we have to use the method of average currents. (The dependence on the power supply voltage is absorbed in the empirical values of $R_{eff,N}$ and $R_{eff,P}$.) Using Eq. (17.9), we write

$$\alpha_n = \frac{2}{\frac{7}{4} - \frac{3 \times 0.35}{0.8} + \left(\frac{0.35}{0.8}\right)^2} = 3.18$$

The value of t_{PHL} can now be found by using Eq. (17.8):

$$t_{PHL} = \frac{3.18 \times 5 \times 10^{-15}}{540 \times 10^{-6} \times 1.5 \times 0.8} = 24.5 \text{ ps}$$

Similarly, the value of $\alpha_p = \alpha_n = 3.18$ can be substituted in Eq. (17.10) to obtain

$$t_{PLH} = \frac{3.18 \times 5 \times 10^{-15}}{(540/5.4) \times 10^{-6} \times 3 \times 0.8} = 66.3 \text{ ps}$$

and t_P can be calculated as

$$t_P = \frac{1}{2}(24.5 + 66.3) = 45.4 \text{ ps}$$

Thus, as expected, reducing V_{DD} has resulted in increased propagation delay by a factor of (45.4/27.8) = 1.63.

Before leaving the subject of propagation delay, we should emphasize that hand analysis using the simple formulas above should *not* be expected to yield precise results. Rather, its value is in obtaining design insight. Precise results can always be obtained using SPICE simulations (see examples in Appendix B and the extensive material on the website). However, it is never a good idea to use simulation without knowing beforehand approximate values of the expected results.

EXERCISES

17.4 For a CMOS inverter fabricated in a 0.18-μm process with $V_{DD} = 1.8$ V, $V_{tn} = -V_{tp} = 0.5$ V, $k_n' = 4k_p' = 300\ \mu A/V^2$, and having $(W/L)_n = 1.5$ and $(W/L)_p = 3$, find t_{PHL}, t_{PLH}, and t_P when the equivalent load capacitance $C = 10$ fF. Use the method of average currents.
Ans. 24.8 ps; 49.6 ps; 37.2 ps

D17.5 For a CMOS inverter fabricated in a 0.13-μm process, use the equivalent-resistances approach to determine $(W/L)_n$ and $(W/L)_p$ so that $t_{PLH} = t_{PHL} = 50$ ps when the effective load capacitance $C =$ 20 fF. Use $R_{eff,N} = 12.5$ kΩ and $R_{eff,P} = 30$ kΩ for $W_{eff}/L_{eff} = 1$.
Ans. 3.5; 8.3

17.1.3 Determining the Equivalent Load Capacitance C

Having determined the propagation delay of the CMOS inverter in terms of the equivalent load capacitance C, it now remains to determine the value of C. For this purpose, a thorough understanding of the various capacitances in a MOS transistor is essential, and we strongly encourage you to review the material in Section 10.1.1.

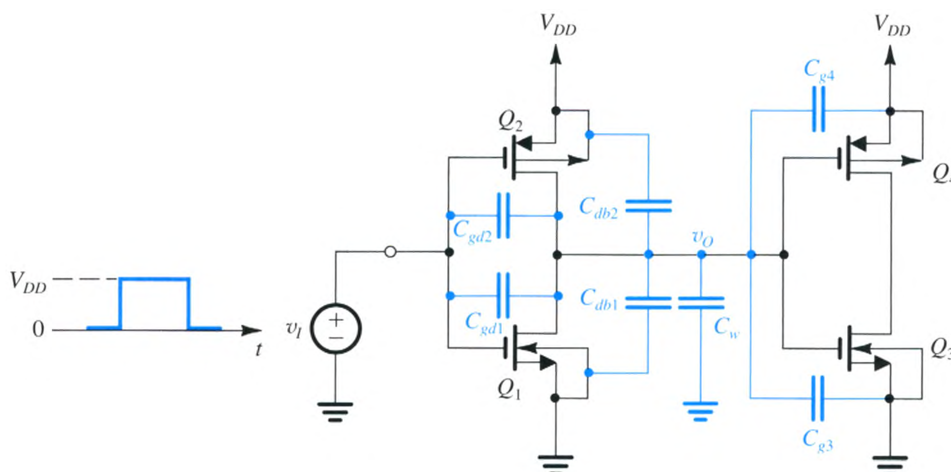


Figure 17.6 Circuit for analyzing the propagation delay of the inverter formed by Q_1 and Q_2, which is driving a similar inverter formed by Q_3 and Q_4.

Figure 17.6 shows the circuit for determining the propagation delay of the CMOS inverter formed by Q_1 and Q_2. Note that we are showing the inverter driving a similar inverter formed by transistors Q_3 and Q_4. This reflects a practical situation and will help us explain how to determine the contribution of a driven inverter to the equivalent capacitance C at the output of the inverter under study (that formed by Q_1 and Q_2).

Shown in Fig. 17.6 are the various transistor capacitances that connect to the output node of the Q_1-Q_2 inverter. Also shown is the **wiring capacitance** C_w, which represents the capacitance of the wire or **interconnect** that connects the output of the Q_1-Q_2 inverter to the input of the Q_3-Q_4 inverter. Interconnect capacitances have become increasingly dominant as the technology has scaled down. In fact, some digital IC designers hold the view that interconnect poses a greater limitation on the speed of operation than the transistors themselves. We will discuss this topic briefly in Section 17.4.

A glance at the circuit in Fig. 17.6 should be sufficient to indicate that a pencil-and-paper analysis is virtually impossible. That, of course, is the reason we opted for the simplification of replacing all these capacitances with an equivalent capacitance C. Before we consider the determination of C, it is useful to observe that during t_{PLH} or t_{PHL}, the output of the first inverter changes from 0 to $V_{DD}/2$ or from V_{DD} to $V_{DD}/2$, respectively. Assuming that the switching threshold of the second inverter is $V_{DD}/2$, it follows that the second inverter remains in the same state during each of our analysis intervals. This observation will have an important bearing on our estimation of the equivalent input capacitance of the second inverter. Let's now consider the contribution of each of the capacitances in Fig. 17.6 to the value of the equivalent load capacitance C:

1. The gate-drain overlap capacitance of Q_1, C_{gd1}, can be replaced by an equivalent capacitance between the output node and ground of $2C_{gd1}$. The factor 2 arises because of the Miller effect (Section 10.2). Specifically, refer to Fig. 17.7 and note that as v_I goes high and v_O goes low by the same amount, the change in voltage across C_{gd1} is twice that amount. Thus the output node sees in effect twice the value of C_{gd1}. The same applies for the gate-drain overlap capacitance of Q_2, C_{gd2}, which is replaced by a capacitance $2C_{gd2}$ between the output node and ground.
2. Each of the drain-body capacitances C_{db1} and C_{db2} has a terminal at a constant voltage. Thus for the purpose of our analysis here, C_{db1} and C_{db2} can be replaced with equal capacitances between the output node and ground. Note, however, that the formula given in Section 10.1.1 for calculating C_{db1} and C_{db2} is a small-signal relationship, whereas the analysis here is obviously a large-signal one. A technique has been developed for finding equivalent large-signal values for C_{db1} and C_{db2} (see Hodges et al., 2004 and Rabaey et al., 2003).
3. Since the second inverter does not switch states, we assume that the input capacitances of Q_3 and Q_4 remain approximately constant and equal to the total gate capacitance $(WLC_{ox}+C_{gsov}+C_{gdov})$. That is, the input capacitance of the load

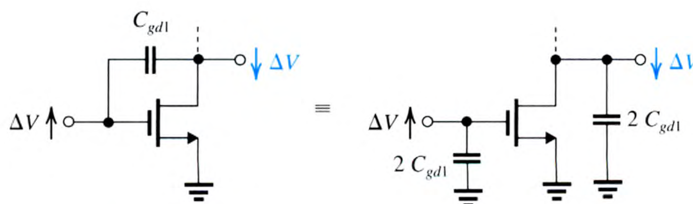


Figure 17.7 The Miller multiplication of the feedback capacitance C_{gd1}.

inverter will be

$$C_{g3} + C_{g4} = (WL)_3 C_{ox} + (WL)_4 C_{ox} + C_{gsov3} + C_{gdov3} + C_{gsov4} + C_{gdov4} \qquad (17.18)$$

4. The last component of C is the wiring capacitance C_w, which simply adds to the value of C.

Thus, the total value of C is given by

$$C = 2C_{gd1} + 2C_{gd2} + C_{db1} + C_{db2} + C_{g3} + C_{g4} + C_w \qquad (17.19)$$

Example 17.3 Determining the Effective Load Capacitance C and the Propagation Delay

Consider a CMOS inverter fabricated in a 28-nm process for which C_{ox} = 34 fF/μm^2, $R_{eff,N}$ = 15 kΩ for $(W/L)_n$ = 1.0, and $R_{eff,P}$ = 20 kΩ for $(W/L)_p$ = 1.0. The W/L ratio of Q_N is 60 nm/28 nm, and that of Q_P is 60 nm/28 nm. The gate-source and gate-drain overlap capacitances are specified to be 0.4 fF/μm of gate width. Further, the effective (large-signal) values of drain-body capacitances are C_{dbn} = 0.03 fF and C_{dbp} = 0.03 fF. The wiring capacitance C_w = 0.5 fF. Find t_{PHL}, t_{PLH}, and t_P when the inverter is driving an identical inverter.

Solution

First, we determine the value of the equivalent capacitance C using Eqs. (17.18) and (17.19):

$$C = 2C_{gd1} + 2C_{gd2} + C_{db1} + C_{db2} + C_{g3} + C_{g4} + C_w$$

where

$$C_{gd1} = 0.4 \times W_n = 0.4 \times 0.06 = 0.024 \text{ fF}$$
$$C_{gd2} = 0.4 \times W_n = 0.4 \times 0.06 = 0.024 \text{ fF}$$
$$C_{db1} = 0.03 \text{ fF}$$
$$C_{db2} = 0.03 \text{ fF}$$
$$C_{g3} = 0.06 \times 0.028 \times 34 + 2 \times 0.4 \times 0.06 = 0.105 \text{ fF}$$
$$C_{g4} = 0.06 \times 0.028 \times 34 + 2 \times 0.4 \times 0.06 = 0.105 \text{ fF}$$
$$C_w = 0.5 \text{ fF}$$

Thus,

$$C = 2 \times 0.024 + 2 \times 0.024 + 0.03 + 0.03 + 0.105 + 0.105 + 0.5 = 0.866 \text{ fF}$$

Next, we use Eqs. (17.14) and (17.15) to determine R_N and R_P:

$$R_N = \frac{15 \times 1.0}{60/28} = 7 \text{ k}\Omega$$

Example 17.3 *continued*

$$R_P = \frac{20 \times 1.0}{60/28} = 9.3 \text{ k}\Omega$$

Now, we can use Eq. (17.12) to determine t_{PHL}

$$t_{PHL} = 0.69 \times 7 \times 10^3 \times 0.866 \times 10^{-15} = 4.18 \text{ ps}$$

and Eq. (17.13) to determine t_{PLH}

$$t_{PLH} = 0.69 \times 9.3 \times 10^3 \times 0.866 \times 10^{-15} = 5.58 \text{ ps}$$

Finally, we determine t_P as

$$t_P = \frac{1}{2}(4.18 + 5.58) = 4.88 \text{ ps.}$$

A common way to report a typical delay for a particular technology is through a **fanout of 4** (FO4) inverter delay, which represents the delay of a single inverter of known dimensions driving a load consisting of four identical inverters, hence where $C_{g3} = 4 \times C_{g1}$ and $C_{g4} = 4 \times C_{g2}$. We can thus modify Eq. (17.19) for an FO4 inverter as follows:

$$C = 2C_{gd1} + 2C_{gd2} + C_{db1} + C_{db2} + 4 \times C_{g1} + 4 \times C_{g2} + C_w \qquad (17.20)$$

EXERCISES

17.6 Consider the inverter specified in Example 17.3 when loaded with an additional 10-fF capacitance. What will the propagation delay become?
Ans. 61.2 ps

17.7 Consider the inverter specified in Example 17.3, including the wire with $C_w = 0.5$ fF. However, in this case the inverter is driving four identical inverters (i.e., FO4). What is the total capacitance C? What is the value of t_P?
Ans. 1.496 fF; 8.43 ps

17.2 Transistor Sizing

We will now address the extremely important design question of selecting appropriate sizes (i.e., L and W/L values) for all transistors in a CMOS logic circuit. We begin with the CMOS inverter and then consider general logic gates.

17.2.1 Inverter Sizing

In this section, we are concerned with selecting appropriate values for the channel length L and the (W/L) ratios for the two transistors Q_N and Q_P in an inverter. Our reasoning can be summarized as follows.

1. To minimize area, the length of all channels is usually made equal to the minimum length permitted by the given technology.
2. In a given inverter, if our interest is strictly to minimize area, $(W/L)_n$ can be selected in the range from 1 to 1.5, which minimizes driving current. However, in more recent technologies it has become common to use slightly higher values, closer to 2 and above.
3. The selection of $(W/L)_n$ relative to $(W/L)_p$ affects the noise margins and t_{PLH}. Both are optimized by matching Q_P and Q_N. Historically, the value of k_n' has significantly exceeded that of k_p', often by a factor of 3 to 5. In those technologies, matching Q_P and Q_N is often wasteful of area and has the potential to increase the effective capacitance C, and although t_{PHL} and t_{PLH} are matched, delay may in fact end up higher than in the case without matching. To avert these problems, selecting $(W/L)_p = (W/L)_n$ is a possibility, and $(W/L)_p = 2(W/L)_n$ is a frequent compromise. In emerging technologies (e.g., the 28-nm process), k_n' and k_p' are closer in value, and it has become more common to use $(W/L)_p = (W/L)_n$.
4. Having settled on an appropriate ratio of $(W/L)_p$ to $(W/L)_n$, we still have to select $(W/L)_n$ to reduce t_P and thus allow higher speeds of operation. Any increase in $(W/L)_n$ and proportionally in $(W/L)_p$ will of course increase area, and hence the inverter contribution to the value of the equivalent capacitance C. To be more precise we express C as the sum of an intrinsic component C_{int} contributed by Q_N and Q_P of the inverter, and an extrinsic component C_{ext} resulting from the wiring and the input capacitance of the driven gates,

$$C = C_{\text{int}} + C_{\text{ext}} \tag{17.21}$$

Increasing $(W/L)_n$ and $(W/L)_p$ of the inverter by a factor S relative to that of a minimum-size inverter for which $C_{\text{int}} = C_{\text{int0}}$ results in

$$C = SC_{\text{int0}} + C_{\text{ext}} \tag{17.22}$$

Now, if we use the equivalent-resistances approach to compute t_P and define an equivalent inverter resistance R_{eq} as

$$R_{\text{eq}} = \frac{1}{2}(R_N + R_P) \tag{17.23}$$

then,

$$t_P = 0.69 R_{\text{eq}} C \tag{17.24}$$

Further, if for the minimum-size inverter R_{eq} is R_{eq0}, increasing $(W/L)_n$ and $(W/L)_p$ by the factor S reduces R_{eq} by the same factor:

$$R_{\text{eq}} = R_{\text{eq0}}/S \tag{17.25}$$

Combining Eqs. (17.24), (17.25), and (17.22), we obtain

$$t_P = 0.69\left(\frac{R_{\text{eq0}}}{S}\right)(SC_{\text{int0}} + C_{\text{ext}})$$

$$t_P = 0.69\left(R_{\text{eq0}}C_{\text{int0}} + \frac{1}{S}R_{\text{eq0}}C_{\text{ext}}\right) \tag{17.26}$$

We can see, then, that scaling the W/L ratios does *not* change the component of t_P caused by the capacitances of Q_N and Q_P. It does, however, reduce the component of t_P that results from capacitances external to the inverter itself. It follows that we can use Eq. (17.26) to decide on a suitable scaling factor S that keeps t_P below a specified maximum value, keeping in mind of course the effect of increasing S on silicon area.

EXERCISE

17.8 For the inverter analyzed in Example 17.3:

(a) Find the intrinsic and extrinsic components of C.

(b) By what factor must $(W/L)_n$ and $(W/L)_p$ be increased to reduce the extrinsic part of t_P by a factor of 2?

(c) Estimate the resulting t_P.

(d) By what factor is the inverter area increased?

Ans. (a) 0.156 fF, 0.71 fF; (b) 2; (c) 2.88 ps; (d) 2

FEDERICO FAGGIN: A PIONEER IN MICROPROCESSOR ELECTRONICS

Holder of a degree in physics from the University of Padua, Federico Faggin first worked for SGS-Fairchild in Italy. In 1968 he relocated to California, joining Fairchild in Palo Alto, where he developed the silicon-gate MOS device that has dominated MOS production ever since. In 1970 he joined Intel, where he led the design and production of the Intel 4004, the world's first commercial single-chip microcomputer, introduced in 1971. This design was based on a four-chip version with separated memory that Ted Hoff had designed in 1969 in response to the request by a Japanese calculator company for a twelve-chip flexible design. The 4004, a 4-bit processor, included 2300 PMOS logic circuits (a long-obsolete logic-circuit form) on a 3 mm × 4 mm die, using a random-logic process created by Faggin. Subsequently, Faggin was responsible for the 8008 at Intel, and the Z80 8-bit microprocessor at Zilog (which he founded in 1974). Later, he went on to cofound several other companies, including Synaptics (in 1986 with Carver Mead and others), which provided touch-sensitive pad and screen designs for the mobile and PC products of many manufacturers.

17.2.2 Transistor Sizing in CMOS Logic Gates

Once a CMOS gate circuit has been generated, the only significant step remaining in the design is to decide on W/L ratios for all devices. These ratios usually are selected to provide the gate with current-driving capability in both directions equal to that of the basic inverter. For the basic inverter design, denote $(W/L)_n = n$ and $(W/L)_p = p$, where n is usually 1 to 3, depending on technology (using larger numbers for newer technologies), and, for a matched design, $p = (\mu_n/\mu_p)\,n$; it should be noted, however, that often $p = 2n$ and for minimum area $p = n$. Thus, we wish to select individual W/L ratios for all transistors in a logic gate so that the PDN should be able to provide a capacitor discharge current *at least* equal to that of an NMOS transistor with $W/L = n$, and the PUN should be able to provide a charging current *at least* equal to that of a PMOS transistor with $W/L = p$. This will guarantee a *worst-case* gate delay equal to that of the basic inverter.[1]

In the preceding description, the idea of "worst case" should be emphasized. It means that in deciding on device sizing, we should find the input combinations that result in the lowest output current and then choose sizes that will make this current equal to that of the basic inverter. Before we consider examples, we need to address the issue of determining the current-driving capability of a circuit consisting of a number of MOS devices. In other words, we need to find the *equivalent W/L ratio* of a network of MOS transistors. To that end, we consider the parallel and series connection of MOSFETs and find the equivalent W/L ratios.

The derivation of the equivalent W/L ratio is based on the fact that the on-resistance of a MOSFET is inversely proportional to W/L (see Eqs. 17.14 and 17.15). Thus, if a number of MOSFETs having ratios of $(W/L)_1$, $(W/L)_2$, . . ., are connected in series, the equivalent series resistance obtained by adding the on-resistances will be

$$\begin{aligned} R_{\text{series}} &= R_{N1} + R_{N2} + \dots \\ &= \frac{\text{constant}}{(W/L)_1} + \frac{\text{constant}}{(W/L)_2} + \dots \\ &= \text{constant}\left[\frac{1}{(W/L)_1} + \frac{1}{(W/L)_2} + \dots\right] \\ &= \frac{\text{constant}}{(W/L)_{\text{eq}}} \end{aligned}$$

resulting in the following expression for $(W/L)_{\text{eq}}$ for transistors connected in series:

$$(W/L)_{\text{eq}} = \frac{1}{\dfrac{1}{(W/L)_1} + \dfrac{1}{(W/L)_2} + \dots} \tag{17.27}$$

Similarly, we can show that the parallel connection of transistors with W/L ratios of $(W/L)_1$, $(W/L)_2$,. . ., results in an equivalent W/L of

$$(W/L)_{\text{eq}} = (W/L)_1 + (W/L)_2 + \dots \tag{17.28}$$

[1]This statement assumes that the total effective capacitance C of the logic gate is the same as that of the inverter. In actual practice, the value of C will be larger for a gate, especially as the fan-in is increased.

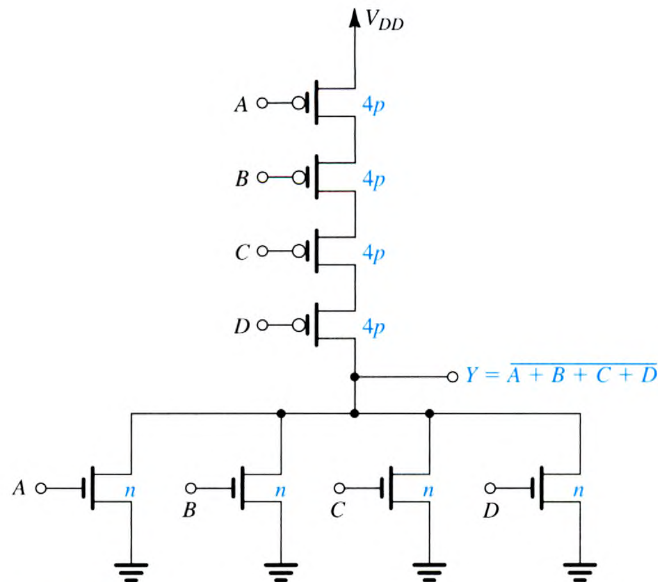


Figure 17.8 Proper transistor sizing for a four-input NOR gate. Note that n and p denote the W/L ratios of Q_N and Q_P, respectively, of the basic inverter.

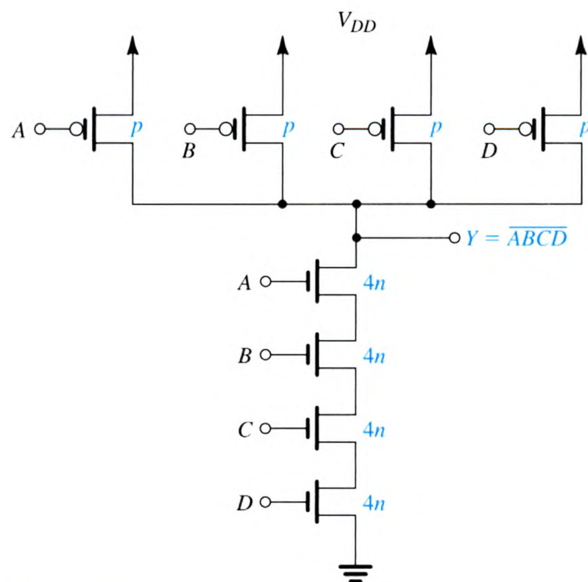


Figure 17.9 Proper transistor sizing for a four-input NAND gate. Note that n and p denote the W/L ratios of Q_N and Q_P, respectively, of the basic inverter.

As an example, two identical MOS transistors with individual W/L ratios of 4 result in an equivalent W/L of 2 when connected in series and of 8 when connected in parallel. We can think about this another way. Connecting MOS transistors in series is equivalent to adding the lengths of their channels while the width does not change; connecting MOS transistors in parallel does not change the channel length but increases the width to the sum of the W's.

As an example of proper sizing, consider the four-input NOR in Fig. 17.8. Here, the worst case (the lowest current) for the PDN is obtained when only one of the NMOS transistors is conducting. We therefore select the *W/L* of each NMOS transistor to be equal to that of the NMOS transistor of the basic inverter, namely, *n*. For the PUN, however, the worst-case situation (and indeed the only case) occurs when all inputs are low and the four series PMOS transistors are conducting. Since the equivalent *W/L* will be one-quarter of that of each PMOS device, we should select the *W/L* ratio of each PMOS transistor to be four times that of Q_p of the basic inverter, that is, $4p$.

As another example, we show in Fig. 17.9 the proper sizing for a four-input NAND gate. Comparison of the NAND and NOR gates in Figs. 17.8 and 17.9 indicates that because p is usually two to three times n, the NOR gate will require much greater area than the NAND gate. For this reason, NAND gates are generally preferred for implementing combinational-logic functions in CMOS.

Example 17.4 Transistor Sizing of a CMOS Gate

Provide transistor *W/L* ratios for the logic circuit shown in Fig. 17.10. Assume that for the basic inverter $n = 2$ and $p = 4$ and that the channel length is 65 nm. Round all values of *W* to the nearest 10 nm.

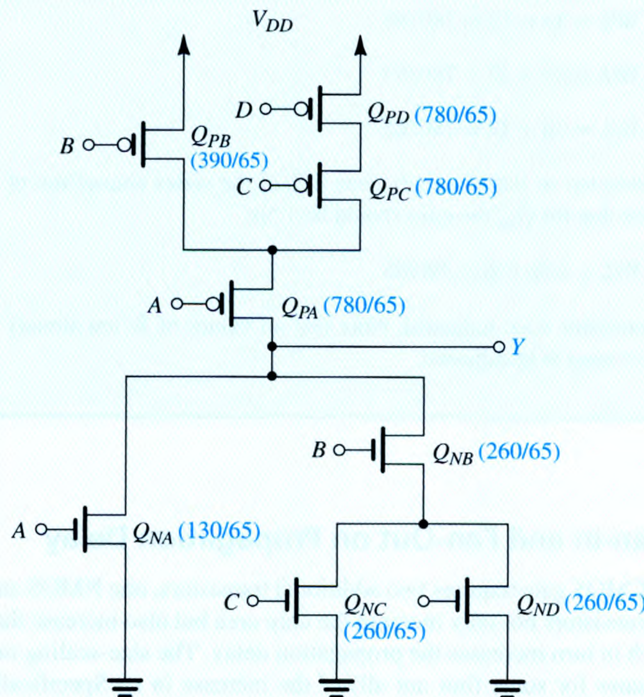


Figure 17.10 Circuit for Example 17.4.

Example 17.4 *continued*

Solution

Refer to Fig. 17.10, and consider the PDN first. Note that the worst case occurs when Q_{NB} is on and either Q_{NC} or Q_{ND} is on. That is, in the worst case, we have two transistors in series. Therefore, we select each of Q_{NB}, Q_{NC}, and Q_{ND} to have twice the width of the n-channel device in the basic inverter:

$$Q_{NB}: W/L = 2n = 4 = 260/65$$

$$Q_{NC}: W/L = 2n = 4 = 260/65$$

$$Q_{ND}: W/L = 2n = 4 = 260/65$$

For transistor Q_{NA}, select W/L to be equal to that of the n-channel device in the basic inverter:

$$Q_{NA}: W/L = n = 2 = 130/65$$

Next, consider the PUN. Here, we see that in the worst case, we have three transistors in series: Q_{PA}, Q_{PC}, and Q_{PD}. Therefore, we select the W/L ratio of each of these to be three times that of Q_P in the basic inverter, that is, $3p$:

$$Q_{PA}: W/L = 3p = 12 = 780/65$$

$$Q_{PC}: W/L = 3p = 12 = 780/65$$

$$Q_{PD}: W/L = 3p = 12 = 780/65$$

Finally, the W/L ratio for Q_{PB} should be selected so that the equivalent W/L of the series connection of Q_{PB} and Q_{PA} should be equal to p. It follows that for Q_{PB} the ratio should be $1.5p$,

$$Q_{PB}: W/L = 1.5p = 6 = 390/65$$

Figure 17.10 shows the circuit with the transistor sizes indicated. Note that all values of W are already rounded to the nearest 10 nm so they do not need to be adjusted.

17.2.3 Effects of Fan-In and Fan-Out on Propagation Delay

Each additional input to a CMOS gate requires two additional transistors, one NMOS and one PMOS. These additional transistors not only increase the chip area but also increase the total capacitance per gate, which in turn increases the propagation delay. The size-scaling method described earlier compensates for some (but not all) of the increase in t_P. Specifically, by increasing device size, we are able to preserve the current-driving capability. However, the capacitance C increases because of both the increased number of inputs and the increase in device size. Thus t_P will still increase with fan-in, a fact that imposes a practical limit on the fan-in of, say, the NAND gate to about 4. If a higher number of inputs is required, then "clever" logic design should be adopted to realize the given Boolean function with gates of

no more than four inputs. This would usually mean an increase in the number of cascaded stages and thus an increase in delay. However, such an increase in delay can be less than the increase due to the large fan-in.

An increase in a gate's fan-out adds directly to its load capacitance and, thus, increases its propagation delay.

Thus although CMOS has many advantages, it does suffer from increased circuit complexity when the fan-in and fan-out are increased, and from the corresponding effects of this complexity on both chip area and propagation delay.

EXERCISES

17.9 For a process technology with $L = 0.18\ \mu\text{m}$, $n = 1.5$, $p = 3$, give the sizes of all transistors in (a) a four-input NOR and (b) a four-input NAND. Also, give the relative areas of the two gates.
Ans.
(a) NMOS devices: $W/L = 0.27/0.18$, PMOS devices: $2.16/0.18$;
(b) NMOS devices: $W/L = 1.08/0.18$, PMOS devices: $0.54/0.18$;
NOR area/NAND area = 1.5

17.10 For the scaled NAND gate in Exercise 17.9, find the ratio of the maximum to minimum current available to (a) charge a load capacitance and (b) discharge a load capacitance.
Ans. (a) 4; (b) 1

17.2.4 Driving a Large Capacitance

In many cases in digital CMOS design, a logic gate must drive a large load capacitance. This might, for example, be due to a long wire on a chip, or to a requirement to drive an off-chip printed-circuit board trace, where the load capacitance can be several hundred times larger than the parasitic capacitances of the driving gate.

Let's investigate how to drive such a large load capacitance without causing the propagation delay to be unacceptably large. Figure 17.11(a) shows the large capacitive load C_L driven by a standard inverter. Note that we have simplified the model of the inverter by assuming that all its capacitances can be lumped into a capacitance C between its input and ground and that it has an effective output resistance R. Connecting C_L directly to the inverter output results in a propagation delay, assuming a ramp input, equal to the time constant τ,

$$t_P = \tau = C_L R \tag{17.29}$$

This propagation delay can be very large.

In an attempt to reduce the propagation delay, we can make the driver inverter large. Such a case is shown in Fig. 17.11(b), where an inverter m times larger than the standard inverter is used. Its output resistance will be R/m, that is, m times lower than that of the standard inverter. As a result, the propagation delay in this case will be

$$\tau = C_L(R/m) = \frac{1}{m}C_L R \tag{17.30}$$

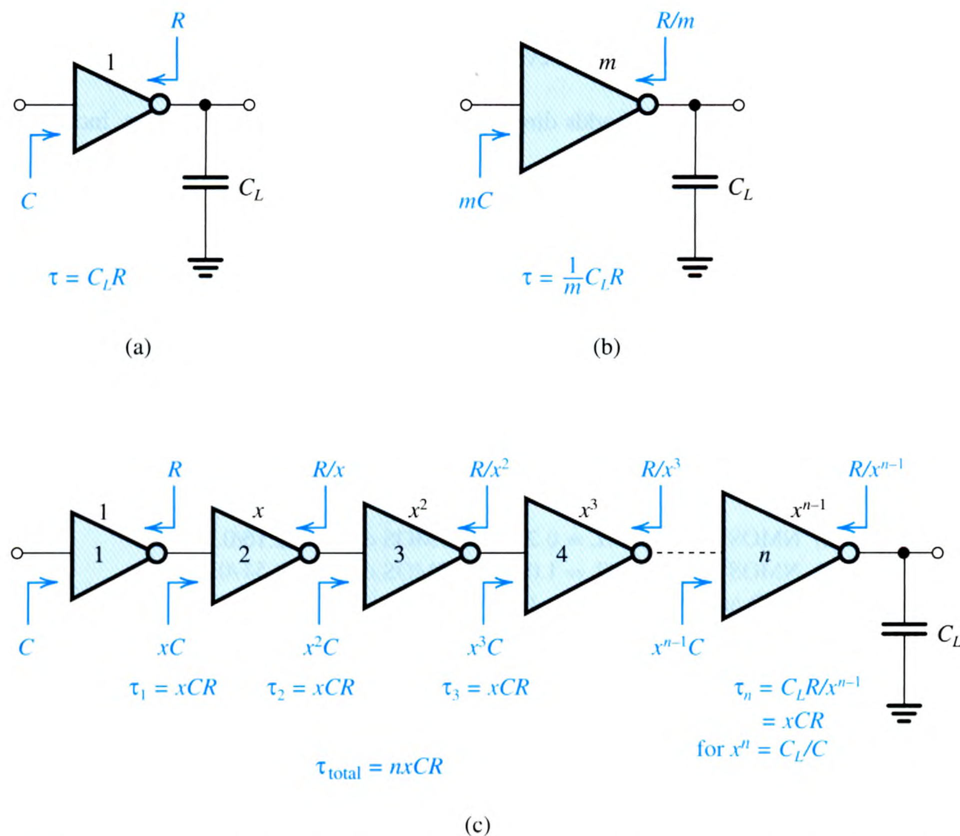


Figure 17.11 Driving a large load capacitance C_L: **(a)** directly; **(b)** by utilizing a large inverter; **(c)** by using a chain of progressively larger inverters.

which as desired has been reduced by a factor m. However, all is not well. Observe that the input capacitance of the large inverter is mC, which can be very large, requiring a large driving inverter to ensure that it does not contribute significantly to lengthening the overall propagation delay. Thus, it appears that we have not solved the problem, but rather shifted the burden to another inverter to drive the input of our large inverter.

The above reasoning leads to the idea of a chain of inverters connected in cascade, as shown in Fig. 17.11(c). Here we have n inverters of progressively larger sizes. In fact, it has been found that the optimum (i.e., lowest overall propagation delay) is obtained when each inverter in the chain is larger than the preceding inverter by the same factor x. Thus if inverter 1 has a unit size, inverter 2 has a size x, inverter 3 has a size x^2, and so on. Figure 17.11(c) shows the effect of inverter size scaling on its input capacitance and its equivalent output resistance. Observe that the delay time associated with the interface between each two succeeding inverters is $\tau = xCR$; that is, each interface contributes equally to the overall delay. This, of course, is a result of the geometric size scaling of the inverters in this chain. It has been shown that minimum delay is obtained if this equality of time constants

extends to the output node, that is, by making

$$\tau_n \equiv C_L\left(\frac{R}{x^{n-1}}\right)$$

equal to xCR, which can be achieved if

$$x^n = \frac{C_L}{C} \tag{17.31}$$

in which case the overall delay becomes

$$t_P = \tau_{\text{total}} = nxCR \tag{17.32}$$

The question of selecting values for x and n remains. First, observe that there is already one condition on their values, namely, that in Eq. (17.31). It can be shown mathematically that the second condition that leads to minimum propagation delay is

$$x = e = 2.718 \tag{17.33}$$

In practice, it has been found that values for x between 2.5 and 4 lead to near-optimum performance (see Hodges et al., 2004).

Example 17.5 Design of an Inverter Chain to Drive a Large Load Capacitance

An inverter whose input capacitance $C = 10$ fF and whose equivalent output resistance $R = 1$ kΩ must ultimately drive a load capacitance $C_L = 1$ pF.

(a) What is the time delay that results if the inverter is connected directly to C_L?
(b) If a driver chain such as that in Fig. 17.11(c) is used, how many inverters n and what size ratio x should you use to minimize the total delay? What is the total path delay achieved?

Solution

(a) $t_P = \tau = 0.69\ C_L R = 10^{-12} \times 10^3 = 0.69$ ns.
(b) The delay is minimized by selecting

$$x = e = 2.718$$

and

$$x^n = \frac{C_L}{C} = \frac{10^{-12}}{10 \times 10^{-15}} = 100$$

which yields

$$n = \frac{\ln 100}{\ln x} = \frac{\ln 100}{\ln e} = 4.6$$

Example 17.5 *continued*

Since we must use an integral number of inverters, we select

$$n = 5$$

and obtain x from

$$x^n = x^5 = \frac{C_L}{C} = 100$$

which yields

$$x = (100)^{1/5} = 2.51$$

The total path delay will be

$$t_P = 0.69\, nxCR$$

$$= 0.69 \times 5 \times 2.51 \times 10 \times 10^{-15} \times 10^3 = 86.6 \text{ ps}$$

which is a reduction in delay by a factor of about 8!

17.3 Power Dissipation

Many of today's integrated circuits are battery powered. Some even rely on "scavenged" energy, thus severely limiting the supply of power. Other high-performance circuits, such as those found at computer server farms, have heat-dissipation limitations. Also, the desire to pack an ever-increasing number of gates on an IC chip (many millions at present) while keeping the power dissipated in the chip to an acceptable limit, has made attending to the power dissipated in a logic-gate circuit of paramount importance. Indeed, minimizing power dissipation in digital ICs is perhaps the most important design challenge today.

In this section, we look at sources of power consumption in digital CMOS circuits and present some metrics that are used in power optimization.

17.3.1 Sources of Power Dissipation

Let us return to the inverter of Fig. 16.17, which dissipates no power when v_I is low and the switch is open. In the other state, however, the power dissipation is approximately V_{DD}^2/R and can be substantial, as we saw in Example 16.2. This power dissipation occurs even if the inverter is not switching and is thus known as **static power dissipation**.

Another inverter we studied earlier (see Fig. 17.12(a)), which is the basis for the CMOS inverter, exhibits no static power dissipation, a definite advantage. Unfortunately, however, another component of power dissipation arises when a capacitance exists between the output node of the inverter and ground. As we have already seen, this is always the case, for the devices that implement the switches have internal capacitances, the wires that connect the inverter output to other circuits have capacitance, and, of course, there is the input capacitance of whatever circuit the inverter is driving. Now, as the inverter is switched from one state to another, current must flow through the switch(es) to charge (and discharge)

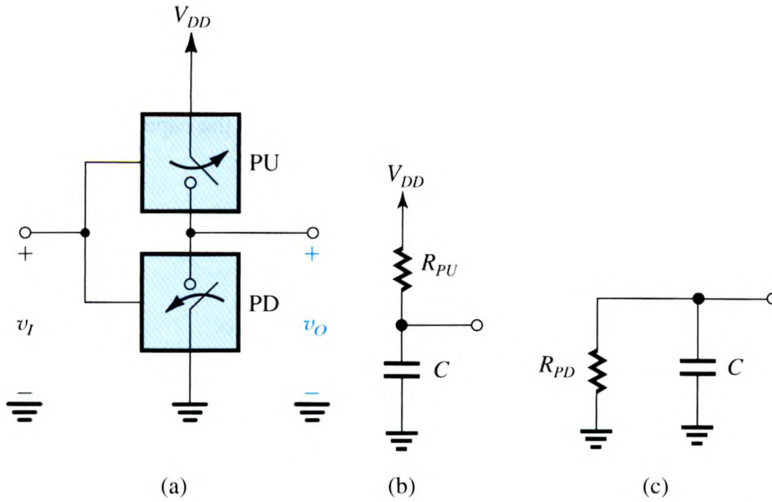


Figure 17.12 **(a)** Switch-level model of a CMOS inverter. **(b)** Equivalent circuit for calculating the dynamic power dissipation when v_I is low. **(c)** Equivalent circuit for calculating the dynamic power dissipation when v_I is high.

the load capacitance. These currents give rise to power dissipation in the switches, called **dynamic power dissipation**.

An expression for the dynamic power dissipation of the inverter of Fig. 17.12(a) can be derived as follows. Consider first the situation when v_I goes low. The pull-down switch P_D turns off and the pull-up switch P_U turns on. In this state, the inverter can be represented by the equivalent circuit shown in Fig. 17.12(b). Capacitor C will charge through the on-resistance of the pull-up switch, and the voltage across C will increase from 0 to V_{DD}. Denoting by $i_D(t)$ the charging current supplied by V_{DD}, we can write for the instantaneous power drawn from V_{DD} the expression

$$p_{DD}(t) = V_{DD} i_D(t)$$

The energy delivered by the power supply to charge the capacitor can be determined by integrating $p_{DD}(t)$ over the charging interval T_c,

$$\begin{aligned} E_{DD} &= \int_0^{T_c} V_{DD} i_D(t) dt \\ &= V_{DD} \int_0^{T_c} i_D(t) dt \\ &= V_{DD} Q \end{aligned}$$

where Q is the charge delivered to the capacitor during the charging interval. Since the initial charge on C was zero,

$$Q = CV_{DD}$$

Thus,

$$E_{DD} = CV_{DD}^2 \tag{17.34}$$

Since at the end of the charging process the energy stored on the capacitor is

$$E_{\text{stored}} = \frac{1}{2}CV_{DD}^2 \tag{17.35}$$

we can find the energy dissipated in the pull-up switch as

$$E_{\text{dissipated}} = E_{DD} - E_{\text{stored}} = \frac{1}{2}CV_{DD}^2 \tag{17.36}$$

This energy is dissipated in the on-resistance of switch PU and is converted to heat.

Next consider the situation when v_I goes high. The pull-up switch PU turns off and the pull-down switch PD turns on. The equivalent circuit in this case is that shown in Fig. 17.12(c). Capacitor C is discharged through the on-resistance of the pull-down switch, and its voltage changes from V_{DD} to 0. At the end of the discharge interval, there will be no energy left on the capacitor. Thus all of the energy initially stored on the capacitor, $\frac{1}{2}CV_{DD}^2$, will be dissipated in the pull-down switch,

$$E_{\text{dissipated}} = \frac{1}{2}CV_{DD}^2 \tag{17.37}$$

This amount of energy is dissipated in the on-resistance of switch PD and is converted to heat.

Thus in each cycle of inverter switching, an amount of energy of $\frac{1}{2}CV_{DD}^2$ is dissipated in the pull-up switch and $\frac{1}{2}CV_{DD}^2$ is dissipated in the pull-down switch, for a total energy loss per cycle of

$$E_{\text{dissipated}}/\text{cycle} = CV_{DD}^2 \tag{17.38}$$

If the inverter is switched at a frequency of f Hz, the dynamic power dissipation of the inverter will be

$$P_{\text{dyn}} = fCV_{DD}^2 \tag{17.39}$$

This is a general expression that does not depend on the inverter circuit details or the values of the on-resistance of the switches.

The expression in Eq. (17.39) indicates that to minimize the dynamic power dissipation, we must strive to reduce the value of C. However, in many cases C is largely determined by the transistors of the inverter itself and cannot be substantially reduced. Another important factor in determining the dynamic power dissipation is the power-supply voltage V_{DD}. Reducing V_{DD} reduces P_{dyn} significantly. This has been a major motivating factor behind the reduction of V_{DD} with every technology generation (see Appendix K). Thus, while the 0.5-μm CMOS process utilized a 5-V power supply, the power-supply voltage used with the 0.13-μm process is only 1.2 V, and that for the 28-nm process is even lower at 0.9 V.

Finally, since P_{dyn} is proportional to the operating frequency f, we may be tempted to reduce P_{dyn} by reducing f. However, this is not a viable proposition in light of the desire to operate digital systems at increasingly higher speeds. These newer chips, however, pack much more circuitry on the chip (several billion transistors) and operate at higher frequencies (microprocessor clock frequencies above 5 GHz are now available). The dynamic power dissipation of such high-density chips can be over 100 W.

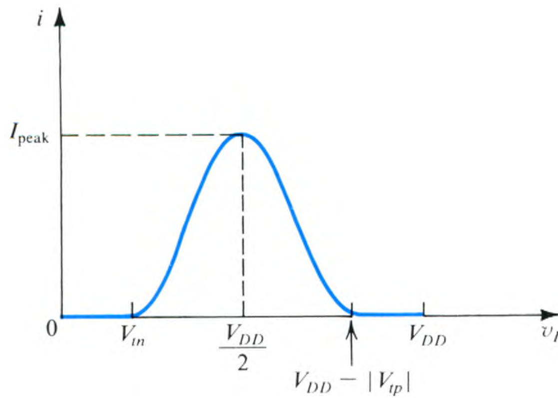


Figure 17.13 The current in the CMOS inverter versus the input voltage.

In addition to the dynamic power dissipation that results from the periodic charging and discharging of the inverter load capacitance, there is another component of power dissipation in the CMOS inverter that results from the current that flows through Q_P and Q_N during every switching event. Figure 17.13 shows this inverter current as a function of the input voltage v_I for a matched inverter. We note that the current peaks at $V_M = V_{DD}/2$. Since at this voltage both Q_N and Q_P operate in saturation, the peak current is given by

$$I_{peak} = \frac{1}{2}\mu_n C_{ox}\left(\frac{W}{L}\right)_n\left(\frac{V_{DD}}{2} - V_{tn}\right)^2 \tag{17.40}$$

The width of the current pulse will depend on the rate of change of v_I with time; the slower the rising edge of the input waveform, the wider the current pulse and the greater the energy drawn from the supply. In general, however, this power component is usually smaller than P_{dyn}.

EXERCISES

17.11 Find the dynamic power dissipation of the inverter analyzed in Example 17.3 when operated at a 1-GHz frequency. Recall that $C = 0.866$ fF and $V_{DD} = 0.9$ V.
Ans. 0.70 μW.

17.12 Find the dynamic power dissipation of a CMOS inverter operated from a 1.2-V supply and having a load capacitance of 80 fF. Let the inverter be switched at 250 MHz.
Ans. 28.8 μW

17.13 A particular inverter circuit initially designed in a 0.5-μm process is fabricated in a 0.13-μm process. Assuming that the capacitance C scales down in proportion to the minimum feature size (more on this in Section 17.4) and that the power supply is reduced from 5 V to 1.2 V, by what factor do you expect the dynamic power dissipation to decrease? Assume that the switching frequency f remains unchanged.
Ans. 66.8

17.3.2 Power–Delay and Energy–Delay Products

The designer is usually interested in high-speed operation (low t_P) combined with low power dissipation. Unfortunately, these two requirements are often in conflict: Generally, if the designer of an inverter attempts to reduce power dissipation by, say, decreasing the supply voltage V_{DD}, or the supply current, or both, the current-driving capability of the inverter decreases. This in turn results in longer times to charge and discharge the load and parasitic capacitances, and thus the propagation delay increases. It follows that a figure of merit for comparing logic-circuit technologies is the **power–delay product** (PDP) of the basic inverter of the given technology, defined as

$$PDP \equiv P_D t_P \tag{17.41}$$

where P_D is the power dissipation of the inverter. Note that the *PDP* is an energy quantity and has the units of joules. The lower the *PDP*, the more effective the inverter and the logic circuits based on the inverter are.

For CMOS logic circuits, the static power dissipation of the inverter is zero,[2] and thus P_D is equal to P_{dyn} and given by Eq. (17.39),

$$P_D = fCV_{DD}^2$$

Thus for the CMOS inverter,

$$PDP = fCV_{DD}^2 t_P \tag{17.42}$$

If the inverter is operated at its theoretical maximum switching speed determined as $f = \frac{1}{2}t_P$, then

$$PDP = \frac{1}{2}CV_{DD}^2 \tag{17.43}$$

From our earlier discussion of dynamic power dissipation we know that $\frac{1}{2}CV_{DD}^2$ is the amount of energy dissipated during each charging or discharging event of the capacitor, that is, for each output transition of the inverter. Thus, the *PDP* has an interesting physical interpretation: *It is the energy consumed by the inverter for each output transition.*

Although the *PDP* is a valuable metric for comparing different technologies for implementing inverters, it is *not* useful as a design parameter for optimizing a given inverter circuit. To appreciate this point, observe that the expression in Eq. (17.43) indicates that the *PDP* can be minimized by reducing V_{DD} as much as possible while, of course, maintaining proper circuit operation. This, however, would not necessarily result in optimal performance, for t_P will increase as V_{DD} is reduced. The problem is that the *PDP* expression in Eq. (17.43) does not in fact have information about t_P. It follows that a better metric can be obtained by multiplying the energy per transition by the propagation delay. We can thus define the **energy–delay product** *EDP* as

$$EDP \equiv \text{Energy per transition} \times t_P$$

$$= \frac{1}{2}CV_{DD}^2 t_P \tag{17.44}$$

[2]The exception to this statement is the power dissipation due to leakage currents and subthreshold conduction in the MOSFETs, discussed in Section 17.4.

EXERCISE

17.14 For the CMOS inverter analyzed in Example 17.3, it was found that $C = 0.866$ fF, $V_{DD} = 0.9$ V, and $t_P = 4.88$ ps. Find the power–delay product when the inverter is operated at its theoretical maximum possible operating frequency. Also find *EDP*.
Ans. 0.35 fJ; 1.7×10^{-27} J · s.

Looking back at Eq. (17.38), we can see that dynamic power consumption is proportional to the switching frequency f. While it may be counterproductive to lower a circuit's overall clock frequency, it may make sense to disable the clock (a technique called **clock gating**) for parts of a circuit that are not actively contributing to computation. This will reduce the number of switching events and dynamic power dissipation.

17.4 Implications of Technology Scaling: Issues in Deep-Submicron Design

The concept of semiconductor technology scaling and the phenomenon known as Moore's law were first discussed in Section 5.4. In this section, we build on that foundation and look at deep-submicron and nanoscale digital circuits, investigating the impact of technology scaling on design practices.

17.4.1 Silicon Area

In addition to minimizing power dissipation and propagation delay, an important objective in the design of digital VLSI circuits is to minimize the silicon area per logic gate. The smaller area makes it possible to fabricate more gates per chip, which has economic and space advantages from a system-design standpoint. Area reduction occurs in three different ways: through advances in processing technology that enable the reduction of the minimum device size, through advances in circuit-design techniques, and through careful chip layout. In this book, our interest lies in circuit design, and we make frequent comments on the relationship between the design of a circuit and its silicon area. As a general rule, the simpler the circuit, the smaller the area required. As we have seen in Section 17.2, the circuit designer has to decide on device sizes. Choosing smaller devices has the obvious advantage of requiring smaller silicon area and at the same time reducing parasitic capacitances and thus increasing speed. Smaller devices, however, have lower current-driving capability, which tends to increase delay. Thus, as in all engineering design problems, there is a trade-off to be made so as to optimize whatever aspect of the design is thought to be critical for the application at hand.

17.4.2 Scaling Implications

In Section 5.4, we introduced the concept of a scaling factor S, by which transistor dimensions (W, L, t_{ox}) and other parameters (V_{DD}, V_t) are scaled from one generation to the next. We saw how these changes have an impact on other metrics, including silicon area, C_{ox}, k_n',

Table 17.1 Implications of Device and Voltage Scaling

	Parameter	Relationship	Scaling Factor
1	W, L, t_{ox}		$1/S$
2	V_{DD}, V_t		$1/S$
3	Area/Device	WL	$1/S^2$
4	C_{ox}	ϵ_{ox}/t_{ox}	S
5	k'_n, k'_p	$\mu_n C_{ox}, \mu_p C_{ox}$	S
6	C_{gate}	WLC_{ox}	$1/S$
7	t_P (intrinsic)	$\alpha C/k'V_{DD}$	$1/S$
8	Energy/Switching cycle (intrinsic)	CV_{DD}^2	$1/S^3$
9	P_{dyn}	$f_{\max} CV_{DD}^2 = \dfrac{CV_{DD}^2}{2t_P}$	$1/S^2$
10	Power density	P_{dyn}/Device area	1

and k'_p. These results were summarized in Table 5.3. We are now in a position to expand the discussion and look at the impact of scaling by a factor S on the speed and power consumption of digital circuits, with results summarized in Table 17.1 (an augmented version of Table 5.3). For easy reference, this table is also included in the Summary Tables supplement found online at www.oup.com/he/sedra-smith8e.

We begin by noting that the component of the inverter propagation delay due to the transistor capacitances (i.e., excluding the wiring capacitance) scales by $1/S$; this very useful result of scaling implies that the circuit can be operated at S times the frequency; that is, the speed of operation increases by a factor S. Equally important, the dynamic power dissipation scales by $1/S^2$. This, of course, is a major motivating factor behind the scaling of V_{DD}. Another motivating factor is the need to keep the electric fields in the MOSFETs within acceptable bounds.

Although the dynamic power dissipation is scaled by $1/S^2$, the power per unit area remains unchanged. Nevertheless, for a number of reasons, as the complexity of digital IC chips continues to increase, so does their power dissipation. Indeed power dissipation has now become the number-one issue in IC design. The problem is exacerbated by the static power dissipation, arising from both subthreshold conduction and diode leakage currents, that plagues deep-submicron CMOS devices. We will discuss this issue, as well as other second-order issues that arise as transistor dimensions keep getting smaller, shortly.

EXERCISES

17.15 By what factor does the power–delay product *PDP* change if an inverter is fabricated in a 0.13-μm technology rather than a 0.25-μm technology? Assume $S \simeq 2$.
Ans. *PDP* decreases by a factor of 8.

17.16 If V_{DD} and V_t are kept constant, which entries in Table 17.1 change and to what value?
Ans. t_P now scales by $1/S^2$; the energy/switching cycle now scales by $1/S$ only; P_{dyn} now scales by S; and the power density now scales by S^3 (a major problem).

17.4.3 Temperature, Voltage, and Process Variations

As we have seen in earlier chapters, temperature variations affect the i–v characteristics of a transistor. Besides affecting the thermal voltage V_T in subthreshold conduction, temperature variations impact the transistor threshold voltage V_t and the mobility μ. These effects, difficult to model in hand calculations, can significantly impact circuit performance, especially when a wide range of temperature environments is expected (we assume you would like your cell phone to work at the beach on a warm day, but also when your car breaks down on the way to a ski resort). Circuit simulators are usually equipped to model temperature effects, and designers should run their simulations at all extreme and expected temperatures.

Supply voltages also vary (e.g., to account for lower battery voltages during extended use), and it is common to require that circuits operate correctly for a range extending to *at least* $V_{DD} \pm 10\%$. Again, this condition can easily be simulated.

Finally, an increasingly important issue in CMOS design is that of process variations. Variations in threshold voltage should be expected, whether at a small scale (transistor-to-transistor), medium scale (die-to-die), or large scale (wafer-to-wafer).

17.4.4 Wiring: The Interconnect

The logic gates on a digital IC chip are connected together by metal wires[3] (see Appendix A). As well, the power-supply V_{DD} and ground are distributed throughout the chip by metal wires. Technology scaling into the deep-submicron range has caused these wires to behave not simply as wires! Specifically, the narrow wires typical of deep-submicron technologies exhibit nonzero resistance. The result is an *IR* drop on the V_{DD} line resulting in somewhat different voltages being delivered to different parts of the chip, as shown in Fig. 17.14. This can compromise the operation of the overall circuit.

Since chips fabricated in deep-submicron technologies can have hundreds of millions of gates, the wire connection between gates can be long. The resulting narrow and long **interconnect** lines have not only nonzero resistance but also capacitance to ground, as shown in Fig. 17.15. The resistance and capacitance of an interconnect line can cause a propagation delay approaching that of the logic gate itself. As well, the capacitance between adjacent wires can cause the signals on one wire to be coupled to the other, which can cause erroneous operation of logic circuits.

[3]These are strips of metal deposited on an insulating surface on top of the chip. In modern digital ICs, several layers of such wiring are utilized.

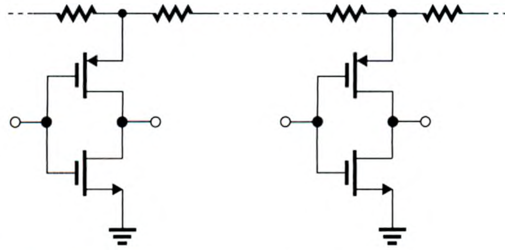


Figure 17.14 The power-supply line in a deep-submicron IC has nonzero resistance. The IR drops along the V_{DD} line cause the voltages delivered to various circuits to differ.

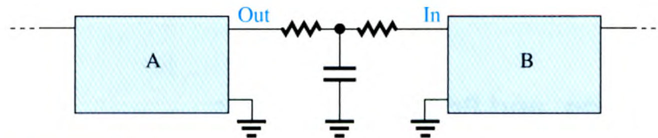


Figure 17.15 The interconnect (wire) between two circuit blocks, A and B, on an IC chip has finite resistance and a capacitance to ground.

In short, the designer of modern deep-submicron digital ICs has to be concerned not only with the logic-circuit design but also with the wiring or interconnect issues. Indeed, so important is the issue of interconnect to digital IC design that advanced textbooks on digital IC design devote entire chapters to this topic (see Rabaey et al., 2003, and Hodges et al., 2004).

17.4.5 Digital Design in Modern Technologies

As transistor dimensions continue to shrink (now into the 10-nm range and below), we start to see deviations from some of the scaling laws presented in Table 17.1 that result from secondary effects. In Section 5.4, we presented two such effects that have a significant impact on digital design: velocity saturation and subthreshold conduction/leakage.

Recall that velocity saturation occurs when the drift velocity of charge carriers in short-channel MOSFETs reaches a limit as the longitudinal electric field established by v_{DS} increases. As a result, drain current i_D does not increase with v_{GS} as rapidly as predicted by the square-law $i - v$ model. Propagation delays get worse as i_D decreases, and therefore, velocity saturation causes delays that are greater than those predicted by the square-law model. Although accurate modeling of velocity saturation is beyond the scope of this textbook, some techniques exist that circuit designers can use without having to resort to complicated pen-and-paper analysis. For instance, accurate device models that account for velocity saturation are embedded in most circuit simulators. Also, when developing a new technology, designers often build test circuits into that technology that are carefully measured to, for example, provide realistic values of R_N and R_P for the equivalent resistance method of Section 17.1.

Another secondary effect with significant impact on digital design is subthreshold conduction. Recall that for $v_{GS} < V_t$, there are small subthreshold drain currents that are proportional to the exponential of v_{GS}/V_T. Although modeling of subthreshold conduction is beyond the scope of this textbook, we note that there is a very strong dependence on temperature due to the thermal voltage V_T. By contributing to static power dissipation,

subthreshold currents represent an important and growing fraction of the total power dissipation in nanoscale circuits. These leakage currents also have consequences for some types of memory circuits, where charge can only be stored on local capacitances for finite amounts of time.

Recently, we have started to see some differentiation in terms of technologies that specifically target either performance (i.e., high speed) or power efficiency (i.e., low power). One way this is done is by deliberately altering the value of V_t to meet design objectives. By lowering V_t and keeping other parameters constant, a transistor can conduct more current, leading to improved speed; on the other hand, when it is turned off, the gate voltage is closer to the threshold point, and subthreshold currents (and hence static power) increase. Conversely, by increasing V_t, speed is poorer, but power consumption can be lowered. There are two basic ways these variations are achieved. (1) Multi-V_t technologies allow designers to select from among predefined high-V_t or low-V_t transistors (some technologies allow even more options). (2) Silicon-on-insulator (SOI) technologies take a different approach: Since the body of each transistor is insulated from others, the body voltage can be individually set to target a desired V_t (recall the body effect), a technique known as *body biasing*. Multi-V_t technologies enable another power-reduction technique known as *power gating*, where high-V_t transistors (i.e., those with low leakage) can be used to turn off the supply to unused portions of a circuit, thus saving on total static power.

As semiconductor technologies continue to evolve and improve, new design challenges will arise, but opportunities for creative circuit design solutions will also present themselves. Below, we highlight some recent innovations that are having a significant impact on digital circuit design.

The FinFET is a recent type of nonplanar MOSFET device that is becoming dominant in digital design for technology nodes below 28 nm. FinFETs are constructed by wrapping a gate around (or partially around) a channel, and they tend to act as better switches. However, because of their nonplanar nature and the manner in which they are fabricated, FinFETs tend to be available only in discretized dimensions, and achieving desired dimensions requires that a designer select an integer number of transistors to connect in parallel, leading to new design practices (albeit some that remind us of the breadboard designs of times past, which used discrete transistors rather than a continuum of W and L values!).

Another development is three-dimensional integration, where several wafers or dies may be stacked on top of each together, with electrical connections formed between layers using through-silicon vias or other connection mechanisms. Vertical integration allows each die to be targeted to a particular application, such as high-performance digital logic closely integrated with Dynamic Random Access Memory (DRAM) technology (which we study in Chapter 18), or perhaps an array of photodiodes for imaging. Such 3D integration is opening up a new and exciting more-than-Moore era.

Summary

- The speed of operation of the inverter is characterized by its propagation delay, t_P. Refer to Fig. 17.3 for the definitions of t_{PLH} and t_{PHL}, and note that $t_P = \frac{1}{2}(t_{PLH} + t_{PHL})$.

- Digital ICs usually use the minimum channel length of the technology available. Thus, for the CMOS inverter, Q_N and Q_P have $L = L_{min}$. If matching is desired, W_p/W_n is selected equal to μ_n/μ_p at the expense of increased area and capacitance. For minimum area, $W_p = W_n$. Also, a frequently used compromise is $W_p = 2W_n$.

- For minimum area, $(W/L)_n$ is selected equal to 1. However, to reduce t_P, especially when a major part of C is extrinsic to the inverter, $(W/L)_n$ and, correspondingly, $(W/L)_p$ can be increased.

- CMOS logic circuits are usually designed to provide equal current-driving capability in both directions. As well, the worst-case values of the pull-up and pull-down

Table 17.2 Summary of Important Speed and Power Characteristics of the CMOS Logic Inverter

Propagation Delay

Using average currents (Fig. 17.4):

$$t_{PHL} \simeq \frac{\alpha_n C}{k_n'(W/L)_n V_{DD}} \text{ where } \alpha_n = \frac{2}{\frac{7}{4} - \frac{3V_{tn}}{V_{DD}} + \left(\frac{V_{tn}}{V_{DD}}\right)^2}$$

$$t_{PLH} \simeq \frac{\alpha_p C}{k_p'(W/L)_p V_{DD}} \text{ where } \alpha_p = \frac{2}{\frac{7}{4} - \frac{3|V_{tp}|}{V_{DD}} + \left(\frac{|V_{tp}|}{V_{DD}}\right)^2}$$

Using equivalent resistances (Fig. 17.5):

$$t_{PHL} = 0.69 R_N C \text{ where } R_N = \frac{R_{eff,N}(W_{eff}/L_{eff})}{(W_n/L_n)}$$

$$t_{PLH} = 0.69 R_P C \text{ where } R_P = \frac{R_{eff,P}(W_{eff}/L_{eff})}{(W_p/L_p)}$$

For a ramp-input signal, $t_{PHL} \simeq R_N C$ and $t_{PLH} \simeq R_P C$.

Power Dissipation

$P_{dyn} = fCV_{DD}^2$

$PDP = P_D \times t_P$

$EDP = PDP \times t_P$

currents are made equal to those of the basic inverter. Transistor sizing is based on this principle and makes use of the equivalent W/L ratios of series and parallel devices [Eqs. (17.27) and (17.28)].

- An important performance parameter of the inverter is the amount of power it dissipates. There are two components of power dissipation: static and dynamic. The first is the result of current flow in either the 0 or 1 state, or both. The second occurs when the inverter is switched and has a capacitor load C. Dynamic power dissipation $P_{dyn} = fCV_{DD}^2$.
- A metric that combines speed of operation and power dissipation is the power-delay product, $PDP = P_D t_P$. The lower the PDP, the more effective the logic-circuit family is. If dynamic power is dominant, such as in CMOS, the delay-power product for an inverter operated at its theoretical maximum switching frequency is $PDP = \frac{1}{2}CV_{DD}^2$, which is the energy drawn from the supply for both a 0-to-1 and a 1-to-0 transition.
- After speed of operation and power dissipation, the silicon area required for an inverter is the third significant metric in digital IC design.
- Predominantly because of its low power dissipation and because of its scalability, CMOS is by far the most dominant technology for digital IC design. This situation is expected to continue for many years to come.
- Refer to Table 17.1 for the implications of scaling MOSFET dimensions, V_{DD}, and V_t by a factor $1/S$. For easy reference, this table is also included in the Summary Tables supplement found online at www.oup.com/he/sedra-smith8e.
- In devices with short channels ($L < 0.25$ μm), velocity saturation occurs. Its effect is that i_D saturates early, and its value is lower than would be predicted by the square-law $i_D - v_{GS}$ characteristic.
- Subthreshold conduction is increasingly becoming an important issue in CMOS circuits, leading to significant static power consumption.
- Table 17.2 summarizes the important speed and power characteristics of the CMOS inverter. For easy reference, this table is also included in the Summary Tables supplement found online at www.oup.com/he/sedra-smith8e.

Computer Simulation Problems

SIM Problems identified by this icon are intended to demonstrate the value of using SPICE simulation to verify hand analysis and design, and to investigate important issues such as allowable signal swing and amplifier nonlinear distortion. Instructions to assist in setting up simulations for all the indicated problems can be found in the corresponding files on the website. Note that if a particular parameter value is not specified in the problem statement, you are to make a reasonable assumption.

Section 17.1: Dynamic Operation of the CMOS Inverter

17.1 For the circuit shown in Fig. P17.1, let switch S open at $t = 0$.

(a) Give the expression for $v_O(t)$.
(b) For $I = 1$ mA and $C = 10$ pF, find the time at which v_O reaches 1 V.

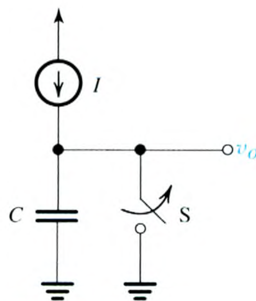


Figure P17.1

17.2 For the circuit in Fig. P17.2, let C be charged to 10 V and switch S closes at $t = 0$.

(a) Give the expression for $v_O(t)$.
(b) For $C = 100$ pF and $R = 1$ kΩ, find t_{PHL} and t_f.

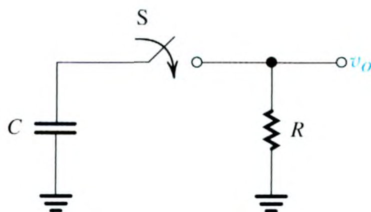


Figure P17.2

17.3 For the inverter circuit in Fig. P17.3, let v_I go from V_{DD} to 0 V at $t = 0$. At $t = 0+$, $v_O = V_{OL}$. Find expressions for V_{OH}, $v_O(t)$, and t_{PLH}. If $R = 10$ kΩ, what is the largest value of C that ensures that t_{PLH} is at most 100 ps?

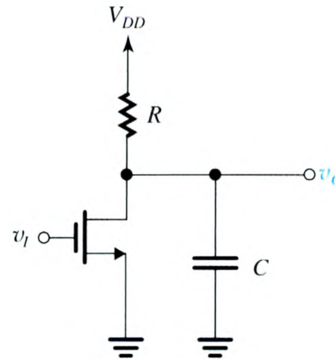


Figure P17.3

17.4 For the inverter of Fig. 16.18(a) with a capacitance C connected between the output and ground, let the on-resistance of PU be 2 kΩ and that of PD be 1 kΩ. If the capacitance $C = 20$ fF, find t_{PLH}, t_{PHL}, and t_P.

17.5 A logic inverter is implemented using the arrangement of Fig. 16.18 with switches having $R_{on} = 2$ kΩ, $V_{DD} = 1.2$ V, and $V_{IL} = V_{IH} = V_{DD}/2$.

(a) Find V_{OL}, V_{OH}, NM_L, and NM_H.
(b) If v_I rises instantaneously from 0 V to +1.2 V and assuming the switches operate instantaneously—that is, at $t = 0$, PU opens and PD closes—find an expression for $v_O(t)$, assuming that a capacitance C is connected between the output node and ground. Hence find the high-to-low propagation delay (t_{PHL}) for $C = 0.1$ pF. Also find t_{THL} (see Fig. 17.3).
(c) Repeat (b) for v_I falling instantaneously from +1.2 V to 0 V. Again assume that PD opens and PU closes instantaneously. Find an expression for $v_O(t)$, and hence find t_{PLH} and t_{TLH}.

17.6 In a particular logic family, the standard inverter, when loaded by a similar circuit, has a propagation delay specified to be 45 ps:

(a) If the current available to charge a load capacitance is half as large as that available to discharge the capacitance, what do you expect t_{PLH} and t_{PHL} to be?

(b) If when an external capacitive load of 0.1 pF is added at the inverter output, its propagation delays increase by 50%, what do you estimate the normal combined capacitance of inverter output and input to be?

(c) If without the additional 0.1 pF load connected, the load inverter is removed and the propagation delays were observed to decrease by 40%, estimate the two components of the capacitance found in (b): that is, the component due to the inverter output and other associated parasitics, and the component due to the input of the load inverter.

***17.7** Consider an inverter for which t_{PLH}, t_{PHL}, t_{TLH}, and t_{THL} are 200 ps, 150 ps, 250 ps, and 100 ps, respectively. The rising and falling edges of the inverter output can be approximated by linear ramps. Also, for simplicity, we define t_{TLH} to be 0% to 100% (rather than 10% to 90%) rise time, and similarly for t_{THL}. Two such inverters are connected in tandem and driven by an ideal input having zero rise and fall times. Calculate the time taken for the output voltage to complete its excursion for (a) a rising input and (b) a falling input. What is the propagation delay for the inverter?

SIM **17.8** For a CMOS inverter fabricated in a 65-nm process with $V_{DD} = 1.0$ V, $V_{tn} = V_{tp} = 0.35$ V, $k_n' = 4k_p' = 500$ A/V^2, and having $(W/L)_n = 2$ and $(W/L)_p = 4$, find t_{PHL}, t_{PLH}, and t_P when the equivalent load capacitance $C = 4$ fF. Use the method of average currents.

17.9 Consider a matched CMOS inverter fabricated in the 65-nm process specified in Problem 17.8. If $C = 10$ fF, use the method of average currents to determine the required (W/L) ratios so that $t_P \leq 25$ ps.

17.10 For the CMOS inverter in Exercise 17.4, use the method of equivalent resistance to determine t_{PHL}, t_{PLH}, and t_P. Use $R_{N,eff} = 12.5$ kΩ and $R_{P,eff} = 30$ kΩ for $W_{eff}/L_{eff} = 1$.

17.11 Use the method of equivalent resistance to determine the propagation delay of a minimum-size inverter, that is, one for which $(W/L)_n = (W/L)_p = 1$, designed in a 0.13-μm technology. The equivalent load capacitance $C = 20$ fF.

17.12 Use the method of equivalent resistance to design an inverter fabricated in a 65-nm technology where transistors with $(W/L) = 1.5$ are found to have $R_{eff,N} = 8$ kΩ and $R_{eff,P} = 24$ kΩ. It is required that for $C = 5$ fF, $t_{PLH} = t_{PHL}$, and $t_P \leq 15$ ps.

17.13 The method of average currents yields smaller values for t_{PHL} and t_{PLH} than those obtained by the method of equivalent resistances. Most of this discrepancy is due to the fact that the formula we derived for I_{av} does not take into account velocity saturation, which reduces the current significantly. Using the results in Example 17.2, by what factor do you estimate the current reduction to be in the NMOS transistor? Since t_{PLH} does not change, what do you conclude about the effect of velocity saturation on the PMOS transistor in this technology?

17.14 Use the method of average currents to estimate t_{PHL}, t_{PLH}, and t_P of a CMOS inverter fabricated in a 65-nm process for which $V_{tn} = |V_{tp}| = 0.35$ V, $V_{DD} = 1$ V, $\mu_n C_{ox} = 470$ μA/V^2, and $\mu_p C_{ox} = 190$ μA/V^2. The inverter has $(W/L)_n = 1.5$ and $(W/L)_p = 3$, and the total capacitance at the inverter output node is 10 fF. Also, find the theoretical maximum frequency at which this inverter can be operated.

17.15 Find the propagation delay for an inverter for which $k_n' = 4k_p' = 500$ μA/V^2 and $(W/L)_n = (W/L)_p = 130$ nm/65 nm, $V_{DD} = 1.0$ V, $V_{tn} = -V_{tp} = 0.35$ V, and the capacitance is roughly 2.5 fF/μm of device width. There is an additional load capacitance of 2 fF. What does t_P become if the design is changed to a matched one? Use the method of average current.

17.16 A matched CMOS inverter fabricated in a process for which $C_{ox} = 25$ fF/μm^2, $\mu_n C_{ox} = 500$ μA/V^2, $\mu_p C_{ox} = 125$ μA/V^2, $V_{tn} = |V_{tp}| = 0.35$ V, and $V_{DD} = 1.0$ V, uses $W_n = 260$ nm and $L_n = L_p = 65$ nm. The overlap capacitance and the effective drain-body capacitance per micrometer of gate width are 0.3 fF and 0.5 fF, respectively. The wiring capacitance is $C_w = 2$ fF. If the inverter is driving another identical inverter, find t_{PHL}, t_{PLH}, and t_P. For how much additional capacitance load does the propagation delay increase by 50%?

Section 17.2: Transistor Sizing

17.17 An inverter whose equivalent load capacitance C is composed of 10 fF, contributed by the inverter transistors, and 30 fF, contributed by the wiring and other external circuitry, has been found to have a propagation delay of 30 ps. By what factor must $(W/L)_n$ and $(W/L)_p$ be increased so as to reduce t_P to 15 ps? By what factor is the inverter area increased?

D *17.18 In this problem, we investigate the effect of the selection of the ratio W_p/W_n on the propagation delay of an inverter driving an identical inverter, as in Fig. 17.6. Assume all transistors have the same L.

(a) Noting that except for C_w each of the capacitances in Eqs. (17.18) and (17.19) is proportional to the width of the relevant transistor, show that C can be expressed as

$$C = C_n\left(1 + \frac{W_p}{W_n}\right) + C_w$$

where C_n is determined by the NMOS transistors.

(b) Using the equivalent resistance method, with $R_{eff,N}(W_{eff}/L_{eff}) = 12.5\ \text{k}\Omega$ and $R_{eff,P}(W_{eff}/L_{eff}) = 30.0\ \text{k}\Omega$, show that for $(W/L)_n = 1$,

$$t_{PHL} = 8.625 \times 10^3 C$$

$$t_{PLH} = \frac{20.7 \times 10^3}{W_p/W_n} C$$

(c) Use the results of (a) and (b) to determine t_P in the case $W_p = W_n$, in terms of C_n and C_w.

(d) Use the results of (a) and (b) to determine t_P in the matched case: that is, when W_p/W_n is selected to yield $t_{PHL} = t_{PLH}$.

(e) Compare the t_P values in (c) and (d) for the two extreme cases:

(i) $C_w = 0$

(ii) $C_w \gg C_n$

What do you conclude about the selection of W_p/W_n?

D 17.19 Consider the CMOS gate shown in Fig. 16.9. Specify W/L ratios for all transistors in terms of the ratios n and p of the basic inverter, such that the worst-case t_{PHL} and t_{PLH} of the gate are equal to those of the basic inverter.

D 17.20 Find appropriate sizes for the transistors used in the exclusive-OR circuit of Fig. 16.10(b). Assume that the basic inverter has $(W/L)_n = 56\ \text{nm}/28\ \text{nm}$ and $(W/L)_p = 84\ \text{nm}/28\ \text{nm}$. What is the total area, including that of the required inverters?

17.21 Consider a four-input CMOS NAND gate for which the transient response is dominated by a fixed-size capacitance between the output node and ground. Compare the values of t_{PLH} and t_{PHL}, obtained when the devices are sized as in Fig. 17.9, to the values obtained when all n-channel devices have $W/L = n$ and all p-channel devices have $W/L = p$.

17.22 Figure P17.22 shows two approaches to realizing the OR function of six input variables. The circuit in Fig. P17.22(b), though it uses additional transistors, has in

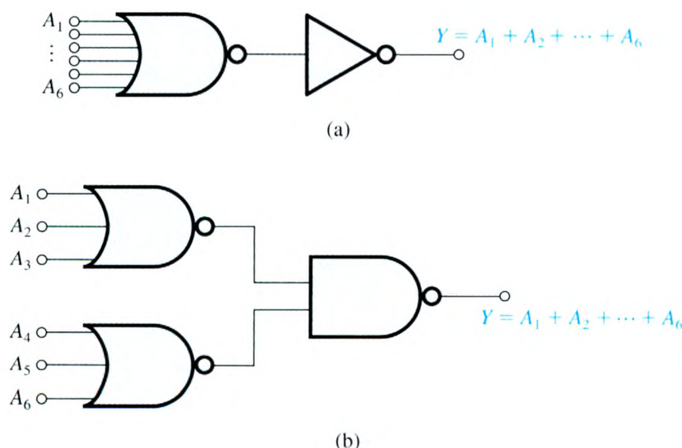


Figure P17.22

fact less total area and lower propagation delay because it uses NOR gates with lower fan-in. Assuming that the transistors in both circuits are properly sized to provide each gate with a current-driving capability equal to that of the basic matched inverter, find the number of transistors and the total area of each circuit. Assume the basic inverter to have a $(W/L)_n$ ratio of 180 nm/90 nm and a $(W/L)_p$ ratio of 360 nm/90 nm.

***17.23** Consider the two-input CMOS NOR gate of Fig. 16.7 whose transistors are properly sized so that the current-driving capability in each direction is equal to that of a matched inverter. For $|V_t| = 0.35$ V and $V_{DD} = 1.0$ V, find the gate threshold in the cases for which (a) input terminal A is connected to ground and (b) the two input terminals are tied together. Neglect the body effect in Q_{PB}.

17.24 A chain of four inverters whose sizes are scaled by a factor x is used to drive a load capacitance $C_L = 1600\,C$, where C is the input capacitance of the standard inverter (which is the first in the chain).

(a) Without increasing the number of inverters in the chain, find the optimum value of x that results in minimizing the overall delay t_P and find the resulting value of t_P in terms of the time constant CR, where R is the output resistance of the standard inverter.
(b) If you are allowed to increase the number of inverters in the chain, what is the number of inverters and the value of x that result in minimizing the total path delay t_P? What is the value of t_P achieved?

17.25 The purpose of this problem is to find the values of n and x that result in minimum path delay t_P for the inverter chain in Fig. 17.11(c).

(a) Show that

$$t_P = \tau_{\text{total}} = (n-1)xRC + \frac{1}{x^{n-1}}RC_L$$

(b) Differentiate the expression for t_P in (a) relative to x and set the derivative to zero. Thus show that the first condition for optimality is

$$x^n = \frac{C_L}{C}$$

(c) Differentiate the expression for t_P in (a) relative to n and set the derivative to zero. Thus show that the second condition for optimality is

$$x^n\left(\frac{C}{C_L}\right) = \ln x$$

(d) Combine the expressions in (b) and (c) to show that the value of x for minimum overall delay is

$$x = e$$

Section 17.3: Power Dissipation

17.26 An IC inverter fabricated in a 65-nm CMOS process is found to have a load capacitance of 3 fF. If the inverter is operated from a 1.2-V power supply, find the energy needed to charge and discharge the load capacitance. If the IC chip has 5 million of these inverters operating at an average switching frequency of 2.5 GHz, what is the power dissipated in the chip? What is the average current drawn from the power supply?

17.27 Consider a standard CMOS logic inverter. Let $V_{DD} = 1$ V, and let a 5-fF capacitance be connected between the output node and ground. Assume there are no other capacitances. If the inverter is switched at the rate of 2.5 GHz, determine the dynamic power dissipation. What is the average current drawn from the dc power supply?

17.28 In a particular logic-circuit technology, operating with a 1.2-V supply, the basic inverter draws (from the supply) a current of 15 μA in one state and 0 μA in the other. When the inverter is switched at the rate of 250 MHz, the average supply current becomes 60 μA. Estimate the equivalent capacitance at the output node of the inverter.

17.29 A collection of logic gates for which the static power dissipation is zero, and the dynamic power dissipation is 10 mW is operating at 1 GHz with a 1.2-V supply. By what fraction could the power dissipation be reduced if operation at 1.0 V were possible? If the frequency of operation is reduced by the same factor as the supply voltage (i.e., 1.0/1.2), what *additional* power can be saved?

17.30 A particular logic gate has t_{PLH} and t_{PHL} of 1.3 ns and 1.2 ns, respectively, and dissipates 0.1 mW with output low and 0.2 mW with output high. Calculate the corresponding delay–power product (under the assumption of a 50% duty-cycle signal and neglecting dynamic power dissipation).

D *17.31 We wish to investigate the design of the inverter shown in Fig. 16.17(a). In particular, we wish to determine the value for R. Selection of a suitable value for R is determined by two considerations: propagation delay and power dissipation.

(a) Show that if v_I changes instantaneously from high to low and assuming that the switch opens instantaneously, the output voltage obtained across a load capacitance C will be

$$v_O(t) = V_{OH} - (V_{OH} - V_{OL})e^{-t/\tau_1}$$

where $\tau_1 = CR$. Hence show that the time required for $v_O(t)$ to reach the 50% point, $\frac{1}{2}(V_{OH} + V_{OL})$, is

$$t_{PLH} = 0.69CR$$

(b) Following a steady state, if v_I goes high and assuming that the switch closes immediately and has the equivalent circuit in Fig. 16.17(c), show that the output falls exponentially according to

$$v_O(t) = V_{OL} + (V_{OH} - V_{OL})e^{-t/\tau_2}$$

where $\tau_2 = C(R \| R_{on}) \simeq CR_{on}$ for $R_{on} \ll R$. Hence show that the time for $v_O(t)$ to reach the 50% point is

$$t_{PHL} = 0.69CR_{on}$$

(c) Use the results of (a) and (b) to obtain the inverter propagation delay, defined as the average of t_{PLH} and t_{PHL} as

$$t_P \simeq 0.35CR \quad \text{for } R_{on} \ll R$$

(d) Show that for an inverter that spends half the time in the logic-0 state and half the time in the logic-1 state, the average static power dissipation is

$$P = \frac{1}{2}\frac{V_{DD}^2}{R}$$

(e) Now that the trade-offs in selecting R should be clear, show that, for $V_{DD} = 1.8$ V and C = 2 pF, to obtain a propagation delay no greater than 1.4 ns and a power dissipation no greater than 2 mW, R should be in a specific range. Find that range and select an appropriate value for R. Then determine the resulting values of t_P and P.

D 17.32 A logic-circuit family with zero static power dissipation normally operates at $V_{DD} = 1.8$ V. To reduce its dynamic power dissipation, operation at 1.2 V is considered. It is found, however, that the currents available to charge and discharge load capacitances also decrease. If current is (a) proportional to V_{DD} or (b) proportional to V_{DD}^2, what reductions in maximum operating frequency do you expect in each case? What fractional change in delay–power product do you expect in each case?

17.33 In this problem, we estimate the CMOS inverter power dissipation resulting from the current pulse that flows in Q_N and Q_P when the input pulse has finite rise and fall times. Refer to Fig. 17.13 and let $V_{tn} = -V_{tp} = 0.4$ V, $V_{DD} = 1.2$ V, and $k_n = k_p = 500\ \mu\text{A/V}^2$. Let the input rising and falling edges be linear ramps with the 0-to-V_{DD} and V_{DD}-to-0 transitions taking 1 ns each. Find I_{peak}. To determine the energy drawn from the supply per transition, assume that the current pulse can be approximated by a triangle with a base corresponding to the time for the rising or falling edge to go from V_t to $V_{DD} - V_t$, and the height equal to I_{peak}. Also, determine the power dissipation that results when the inverter is switched at 100 MHz.

Section 17.4: Implications of Technology Scaling: Issues in Deep-Submicron Design

17.34 Consider the scaling from a 0.13-μm process to a 65-nm process.

(a) Assuming V_{DD} and V_t are scaled by the same factor as the device dimensions ($S = 2$), find the factor by which t_P, the maximum operating speed, P_{dyn}, power density, and PDP decrease (or increase)?

(b) Repeat (a) for the situation in which V_{DD} and V_t remain unchanged.

17.35 An NMOS transistor with $k_n = 0.4$ mA/V^2 and a nominal V_{tn} of 0.4 V is to operate in saturation at $I_D = 0.2$ mA.

(a) If V_{tn} can vary by as much as ±10%, what is the expected range of I_D obtained?

(b) If the transistor is used to discharge a 100-fF load capacitance, what is the expected variation in delay time, assuming that the output voltage is to change by 0.1 V?

CHAPTER 18

Memory and Clocking Circuits

IN THIS CHAPTER YOU WILL LEARN

- How a transmission gate, which is realized using one NMOS and one PMOS transistor, can act as a very effective switch.
- How the basic bistable circuit, the latch, is realized by connecting two inverters in a positive-feedback loop.
- How to augment the latch to obtain different types of flip-flops that are useful building blocks for digital systems.
- How CMOS is particularly well suited for the efficient implementation of a particular type of flip-flop, the D flip-flop.
- How to analyze and design a six-transistor circuit that is used almost universally to implement the storage cell in static random-access memory (SRAM), and a one-transistor circuit that is equally universal in the implementation of the storage cell in dynamic random-access memory (DRAM).
- How some special-purpose circuits can provide required timing information for sequential operation.
- Interesting circuit techniques for accessing a particular storage cell in a memory chip and for amplifying the signal readout from the cell.

Introduction

The logic circuits studied in Chapters 16 and 17 are called **combinational circuits**. Their output depends only on the present value of the input. Thus these circuits do *not* have memory. *Memory* is a very important part of digital systems. Its availability in digital computers allows for storing programs and data. Furthermore, it is important for temporary storage of the output produced by a combinational circuit for use at a later time in the operation of a digital system.

Logic circuits that incorporate memory are called **sequential circuits**; that is, their output depends not only on the present value of the input but also on the input's previous values. Such circuits require a timing generator (a *clock*) for their operation.

There are basically two approaches for providing memory to a digital circuit. The first relies on the application of positive feedback, which can be arranged to provide a circuit with two stable states. Such a *bistable* circuit can then be used to store one bit of information: One stable state would correspond to a stored 0, and the other to a stored 1. A bistable circuit can remain in either state indefinitely, and thus it belongs to the category of *static sequential circuits*. The other approach to realizing memory uses the storage of charge on a capacitor: When the capacitor is charged, it would be regarded as storing a 1; when it is discharged, it would be storing a 0. Since the inevitable leakage effects will cause the capacitor to discharge, such a form of memory requires the periodic recharging of the capacitor, a process known as *refresh*. Thus, memory based on charge storage is known as *dynamic memory* and the corresponding sequential circuits as *dynamic sequential circuits*.

This chapter concerns the study of memory circuits. We begin in Section 18.1 with a very simple circuit with only two transistors—one NMOS and one PMOS—called a **transmission gate**, which will become a useful building block in the design of more complicated memory elements such as latches, flip-flops, and memory cells. Then, Section 18.2 deals with the basic bistable circuit—the latch—and its application in flip-flops, an important class of building blocks for sequential digital systems. After a brief review of memory-chip organization and nomenclature, Section 18.3 studies memory cell circuits for static memory (SRAM) and dynamic memory (DRAM). The chapter concludes in Section 18.4 with a discussion of several peripheral and special-purpose circuits that are necessary for the proper operation of memories and other sequential circuits, including ring oscillators, sense amplifiers, and decoders.

18.1 The Transmission Gate

A conceptually simple approach for implementing logic functions utilizes series and parallel combinations of switches that are controlled by input logic variables to connect the input and output nodes (see Fig. 18.1). Each of the switches can be implemented either by a single NMOS transistor [Fig. 18.2(a)] or by a pair of complementary MOS transistors connected in what is known as the **CMOS transmission-gate** configuration [Fig. 18.2(b)]. The result is a simple form of logic circuit that is particularly suited for some special logic functions and is frequently used with standard CMOS logic to implement such functions efficiently: that is, with fewer transistors than is possible with CMOS alone.

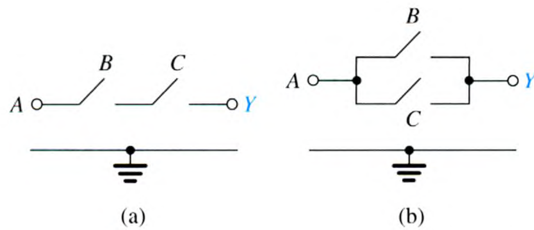


Figure 18.1 Conceptual pass-transistor logic gates. **(a)** Two switches, controlled by the input variables B and C, when connected in series in the path between the input node to which an input variable A is applied and the output node (with an implied load to ground) realize the function $Y = ABC$. **(b)** When the two switches are connected in parallel, the function realized is $Y = A(B + C)$.

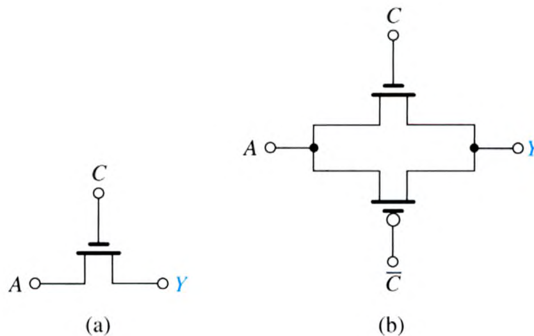


Figure 18.2 Two possible implementations of a voltage-controlled switch connecting nodes A and Y: **(a)** single NMOS transistor and **(b)** CMOS transmission gate.

Because this form of logic utilizes MOS transistors in the series path from input to output, to *pass* or block signal transmission, it is known as *pass-transistor logic* (PTL). As mentioned earlier, CMOS transmission gates are frequently used to implement the switches, giving this logic-circuit form the alternative name *transmission-gate logic*. The terms are used interchangeably independent of the actual implementation of the switches.

18.1.1 Operation with NMOS Transistors as Switches

Implementing the switches in a PTL circuit with single NMOS transistors results in a simple circuit with small area and small node capacitances. These advantages, however, are obtained at the expense of serious shortcomings in both the static characteristics and the dynamic performance of the resulting circuits. To illustrate, consider the circuit shown in Fig. 18.3, where an NMOS transistor Q is used to implement a switch connecting an input node with voltage v_I and an output node. The total capacitance between the output node and ground is represented by capacitor C. The switch is shown in the closed state with the control signal applied to its gate being high at V_{DD}. We wish to analyze the operation of the circuit as the input voltage v_I goes high (to V_{DD}) at time $t = 0$. We assume that initially the output voltage v_O is zero and capacitor C is fully discharged.[1]

[1]Although the MOS transistor is symmetric and its drain and source are interchangeable, it is always useful to know which terminal is functioning as the source and which as the drain. The terminal with the higher voltage in an NMOS transistor is the drain. The opposite is true for the PMOS transistor.

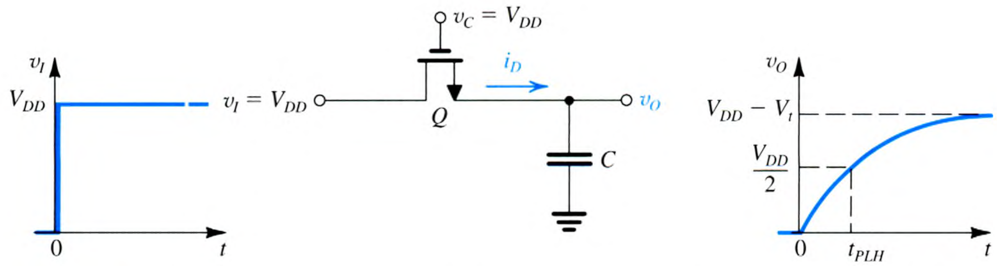


Figure 18.3 Operation of the NMOS transistor as a switch in the implementation of PTL circuits. This analysis is for the case with the switch closed (v_C is high) and the input going high ($v_I = V_{DD}$).

When v_I goes high, the transistor operates in the saturation mode and delivers a current i_D to charge the capacitor,

$$i_D = \frac{1}{2}k_n(V_{DD} - v_O - V_t)^2 \tag{18.1}$$

where $k_n = k_n'(W/L)$, and V_t is determined by the body effect since the source is at a voltage v_O relative to the body (which, though not shown, is connected to ground); thus (see Eq. 5.30),

$$V_t = V_{t0} + \gamma\left(\sqrt{v_O + 2\phi_f} - \sqrt{2\phi_f}\right) \tag{18.2}$$

Thus, initially (at $t = 0$), $V_t = V_{t0}$ and the current i_D is relatively large. However, as C charges up and v_O rises, V_t increases (Eq. 18.2) and i_D decreases. The latter effect is due to the increase both in v_O and in V_t. It follows that the process of charging the capacitor will be relatively slow. More seriously, observe from Eq. (18.1) that i_D reduces to zero when v_O reaches (V_{DD}–V_t). Thus the high output voltage (V_{OH}) will *not* be equal to V_{DD}; rather, it will be lower by V_t, and to make matters worse, the value of V_t can be as high as 1.5 to 2 times V_{t0}!

The propagation delay t_{PLH} of the PTL gate of Fig. 18.3 can be determined as the time for v_O to reach $V_{DD}/2$. This can be calculated using techniques similar to those employed in the analysis of the CMOS inverter in Section 17.1, as will be illustrated shortly in an example.

Figure 18.4 shows the NMOS switch circuit when v_I is brought down to 0 V. We assume that initially $v_O = V_{DD}$. Thus at $t = 0+$, the transistor conducts and operates in the saturation region,

$$i_D = \frac{1}{2}k_n(V_{DD} - V_t)^2 \tag{18.3}$$

where we note that since the source is now at 0 V (note that the drain and source have interchanged roles), there will be no body effect, and V_t remains constant at V_{t0}. As C discharges, v_O decreases and the transistor enters the triode region at $v_O = V_{DD} - V_t$. Nevertheless, the capacitor discharge continues until C is fully discharged and $v_O = 0$. Thus, the NMOS transistor provides $V_{OL} = 0$, or a "good 0." Again, the propagation delay t_{PHL} can be determined using usual techniques, as illustrated by the following example.

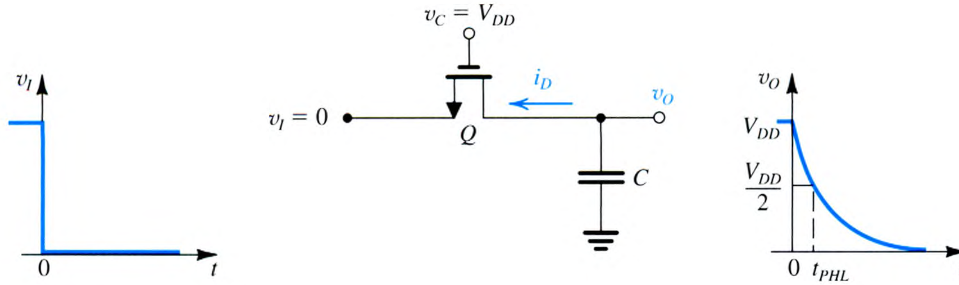


Figure 18.4 Operation of the NMOS switch as the input goes low ($v_I = 0$ V). Note that the drain of an NMOS transistor is always higher in voltage than the source; correspondingly, the drain and source terminals interchange roles in comparison to the circuit in Fig. 18.3.

Example 18.1

Consider the NMOS transistor switch in the circuits of Figs. 18.3 and 18.4 to be fabricated in a technology for which $\mu_n C_{ox} = 450\ \mu\text{A/V}^2$, $\mu_p C_{ox} = 150\ \mu\text{A/V}^2$, $V_{t0} = 0.35$ V, $\gamma = 0.25\ \text{V}^{1/2}$, $2\phi_f = 1.0$ V, and $V_{DD} =$ 1 V. Let the transistor have $(W/L) = 130$ nm/65 nm, and assume that the total capacitance between the output node and ground is $C = 5$ fF.

(a) For the case with v_I high (Fig. 18.3), find V_{OH}.
(b) If the output feeds a CMOS inverter whose $(W/L)_p = 3(W/L)_n = 390$ nm/65 nm, find the static current of the inverter and its power dissipation when its input is at the value found in (a). Also, find the inverter output voltage.
(c) Find t_{PLH}.
(d) For the case where v_I going low (Fig. 18.4), find t_{PHL}.
(e) Find t_P.

Solution

(a) Refer to Fig. 18.3. If V_{OH} is the value of v_O at which Q stops conducting,

$$V_{OH} = V_{DD} - V_t$$

where V_t is the value of the threshold voltage at a source-body reverse bias equal to V_{OH}. Using Eq. (18.2), we have

$$V_t = V_{t0} + \gamma\left(\sqrt{V_{OH} + 2\phi_f} - \sqrt{2\phi_f}\right) = V_{t0} + \gamma\left(\sqrt{V_{DD} - V_t + 2\phi_f} - \sqrt{2\phi_f}\right)$$

Substituting $V_{t0} = 0.35$, $\gamma = 0.25$, $V_{DD} = 1$, and $2\phi_f = 1$, we obtain a quadratic equation in V_t whose solution yields $V_t = 0.41$ V. Thus, $V_{OH} = 0.59$ V. Note that this represents a significant loss in signal amplitude.

(b) The load inverter will have an input signal of 0.59 V. Thus, its Q_P will conduct a current

$$i_{DP} = \frac{1}{2} \times 150 \times \frac{390}{65}(1 - 0.59 - 0.35)^2 = 1.62\ \mu\text{A}$$

Example 18.1 *continued*

where we have assumed Q_P to be operating in saturation, as we still expect v_O of the inverter to be close to 0 V. Thus, the static power dissipation of the inverter will be

$$P_D = V_{DD} i_{DP} = 1 \times 1.62 = 1.62\ \mu\text{W}$$

The output voltage of the inverter can be found by noting that Q_N will be operating in the triode region. Equating its current to that of Q_P (i.e., 1.62 μA) enables us to determine the output voltage to be 7.6 mV.

(c) To determine t_{PLH} , refer to Fig. 18.3. We need to find the current i_D at $t = 0$ (where $v_O = 0$, $V_t = V_{t0} =$ 0.35 V) and at $t = t_{PLH}$ (where $v_O = 0.5$ V, V_t to be determined), as follows:

$$i_D(0) = \frac{1}{2} \times 450 \times \frac{130}{65} \times (1 - 0.35)^2 = 190\ \mu\text{A}$$

$$V_t(\text{at } v_O = 0.5\text{ V}) = 0.35 + 0.25\left(\sqrt{0.5 + 1.0} - \sqrt{1.0}\right) = 0.41\text{ V}$$

$$i_D(t_{PLH}) = \frac{1}{2} \times 450 \times \frac{130}{65} \times (1 - 0.5 - 0.41)^2 = 3.6\ \mu\text{A}$$

We can now compute the average discharge current as

$$i_D|_{av} = \frac{190 + 3.6}{2} = 96.8\ \mu\text{A}$$

and t_{PLH} can be found as

$$t_{PLH} = \frac{C(V_{DD}/2)}{i_D|_{av}} = \frac{5 \times 10^{-15} \times 0.5}{96.8 \times 10^{-6}} = 26\text{ ps}$$

(d) Refer to the circuit in Fig. 18.4. Observe that, here, V_t remains constant at $V_{t0} = 0.35$ V. At $t = 0$, Q will be operating in saturation, and the drain current will be

$$i_D(0) = \frac{1}{2} \times 450 \times \frac{130}{65}(1 - 0.35)^2 = 190\ \mu\text{A}$$

At $t = t_{PHL}$, Q will be operating in the triode region, and thus

$$i_D(t_{PHL}) = 450 \times \frac{130}{65}\left[(1 - 0.35) \times 0.5 - \frac{1}{2} \times 0.5^2\right] = 180\ \mu\text{A}$$

Thus, the average discharge current is given by

$$i_D|_{av} = \frac{190 + 180}{2} = 185\ \mu\text{A}$$

and t_{PHL} can be determined as

$$t_{PHL} = \frac{5 \times 10^{-15} \times 0.5}{185 \times 10^{-6}} = 14\text{ ps}$$

(e) $t_P = \frac{1}{2}(t_{PLH} + t_{PHL}) = \frac{1}{2}(26 + 14) = 20\text{ ps}$

EXERCISE

18.1 Let the NMOS transistor switch in Fig. 18.3 be fabricated in a 0.18-μm CMOS process for which $V_{t0} = 0.5$ V, $\gamma = 0.3$ V$^{1/2}$, $2\phi_f = 0.85$ V, and $V_{DD} = 1.8$ V. Find V_{OH}.
Ans. 1.15 V

18.1.2 Restoring the Value of V_{OH} to V_{DD}

Example 18.1 highlights the problem of signal-level loss and its unfavorable effect on the operation of the succeeding CMOS inverter. Some rather ingenious techniques have been developed to restore the output level to V_{DD}.

Consider the approach illustrated in Fig. 18.5. Here, Q_1 is a pass-transistor controlled by input B. The output node of the pass transistor is connected to the input of a standard CMOS inverter formed by Q_N and Q_P. A PMOS transistor Q_R, whose gate is controlled by the output voltage of the inverter, v_{O2}, has been added to the circuit. Observe that in the event that the output of the PTL gate, v_{O1}, is low (at ground), v_{O2} will be high (at V_{DD}), and Q_R will be off. On the other hand, if v_{O1} is high but not quite equal to V_{DD}, the output of the inverter will be low (as it should be) and Q_R will turn on, supplying a current to charge C up to V_{DD}. This process will stop when $v_{O1} = V_{DD}$, that is, when the output voltage has been restored to its proper level. The "level-restoring" function performed by Q_R is frequently used in MOS digital circuit design. Although the description of operation is relatively straightforward, the addition of Q_R closes a "positive-feedback" loop around the CMOS inverter, and thus operation is more involved than it appears, especially during transients. Selecting a W/L ratio for Q_R is also complicated, although normally k_r is selected to be much lower than k_n (say a third or a fifth as large). Intuitively, this is appealing, for it implies that Q_R will not play a major role in circuit operation, apart from restoring the level of V_{OH} to V_{DD}, as explained above. Transistor Q_R is said to be a "weak PMOS transistor." See Problem 18.9.

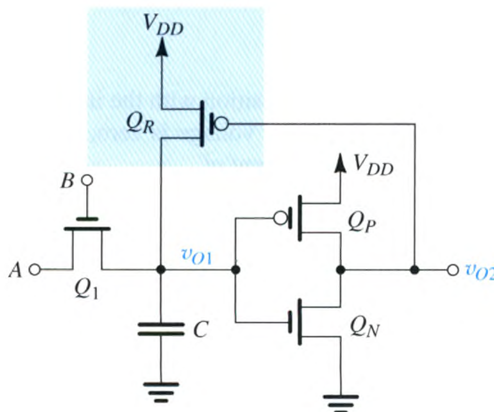


Figure 18.5 The use of transistor Q_R, connected in a feedback loop around the CMOS inverter, to restore the V_{OH} level, produced by Q_1, to V_{DD}.

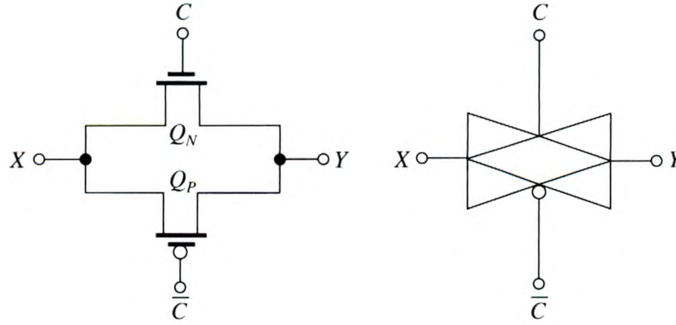


Figure 18.6 The CMOS transmission gate and its circuit symbol.

18.1.3 The Use of CMOS Transmission Gates as Switches

We get great improvements in static and dynamic performance when we implement the switches with CMOS transmission gates. The transmission gate uses a pair of complementary transistors connected in parallel. It acts as an excellent switch, providing bidirectional current flow, and it exhibits an on-resistance that remains almost constant for wide ranges of input voltage. These characteristics make the transmission gate not only an excellent switch in digital applications but also an excellent analog switch in such applications as data converters and switched-capacitor filters (Chapter 14).

Before we analyze the transmission-gate circuit, it is useful to reflect on its origin. Recall that an NMOS transistor transmits the 0-V level to the output perfectly and thus produces a "good 0." It has difficulty, however, in passing the V_{DD} level, with the result that $V_{OH} = V_{DD} - V_t$ (a "poor 1"). It can be shown (see Problems 18.1 and 18.7) that a PMOS transistor does exactly the opposite; that is, it passes the V_{DD} level perfectly and thus produces a "good 1" but has trouble passing the 0-V level, thus producing a "poor 0." It is natural therefore to think that placing an NMOS and a PMOS transistor in parallel would produce good results in both the 0 and 1 cases.

Another way to describe the performance of the two transistor types is that the NMOS is good at pulling the output down to 0 V, while the PMOS is good at pulling the output up to V_{DD}. Interestingly, these are also the roles they play in the standard CMOS inverter.

Figure 18.6 shows the transmission gate together with its frequently used circuit symbol. The transmission gate is a bilateral switch that results in $v_Y = v_X$ when v_C is high (V_{DD}). In terms of logic variables, its function is described by

$$Y = X \quad \text{if} \quad C = 1$$

Figure 18.7(a) shows the transmission-gate switch in the "on" position with the input, v_I, rising to V_{DD} at $t = 0$. Assuming, as before, that initially the output voltage is zero, we see that Q_N will be operating in saturation and providing a charging current of

$$i_{DN} = \frac{1}{2}k_n(V_{DD} - v_O - V_{tn})^2 \tag{18.4}$$

where, as in the case of the single NMOS switch, V_{tn} is determined by the body effect,

$$V_{tn} = V_{t0} + \gamma\left(\sqrt{v_O + 2\phi_f} - \sqrt{2\phi_f}\right) \tag{18.5}$$

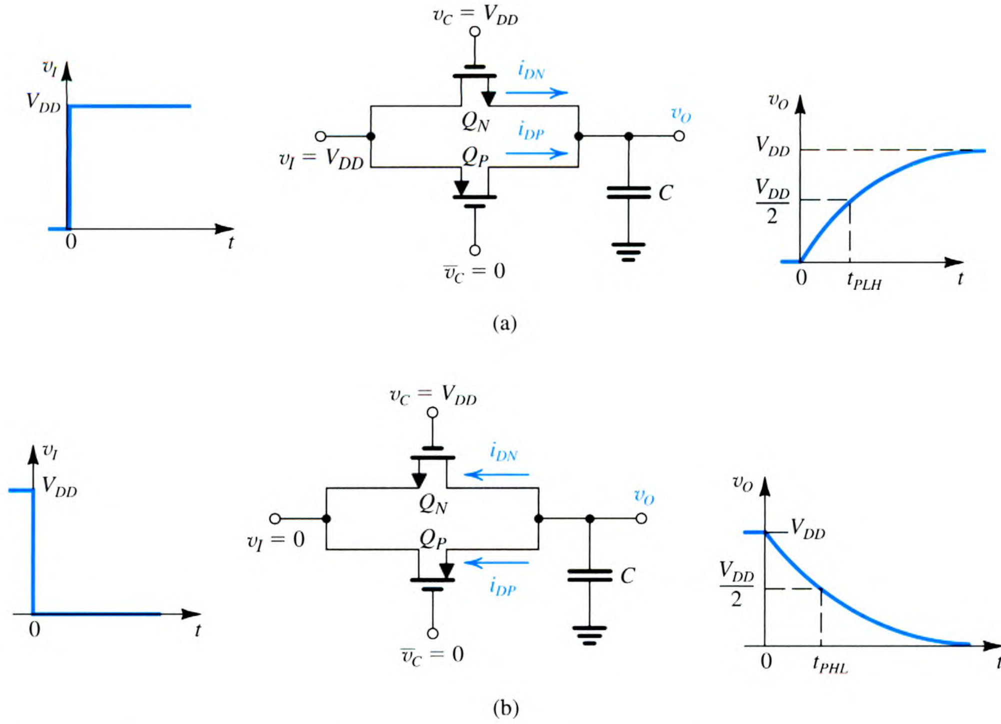


Figure 18.7 Operation of the transmission gate as a switch with **(a)** v_I high and **(b)** v_I low.

Transistor Q_N will conduct a diminishing current that reduces to zero at $v_O = V_{DD} - V_{tn}$. Observe, however, that Q_P operates with $V_{SG} = V_{DD}$ and is initially in saturation,

$$i_{DP} = \frac{1}{2}k_p\left(V_{DD} - |V_{tp}|\right)^2 \tag{18.6}$$

where, since the body of Q_P is connected to V_{DD}, $|V_{tp}|$ remains constant at the value V_{t0}, assumed to be the same value as for the n-channel device. The total capacitor-charging current is the sum of i_{DN} and i_{DP}. Now, Q_P will enter the triode region at $v_O = |V_{tp}|$, but will continue to conduct until C is fully charged and $v_O = V_{OH} = V_{DD}$. Thus, the p-channel device will provide the gate with a "good 1." The value of t_{PLH} can be calculated using usual techniques, where we expect that as a result of the additional current available from the PMOS device, for the same value of C, t_{PLH} will be lower than in the case of the single NMOS switch. Note, however, that adding the PMOS transistor increases the value of C.

When v_I goes low, as shown in Fig. 18.7(b), Q_N and Q_P interchange roles. Analysis of the circuit in Fig. 18.7(b) will indicate that Q_P will cease conduction when v_O falls to $|V_{tp}|$, where $|V_{tp}|$ is given by

$$|V_{tp}| = V_{t0} + \gamma\left[\sqrt{V_{DD} - v_O + 2\phi_f} - \sqrt{2\phi_f}\right] \tag{18.7}$$

Transistor Q_N, however, continues to conduct until C is fully discharged and $v_O = V_{OL} = 0$ V, a "good 0."

We conclude that transmission gates provide far superior performance, both static and dynamic, than is possible with single NMOS switches. The price paid is increased circuit complexity, area, and capacitance.

EXERCISE

18.2 The transmission gate of Fig. 18.7 is fabricated in a CMOS process technology for which $k_n' = 450\ \mu A/V^2$, $k_p' = 150\ \mu A/V^2$, $V_{tn} = |V_{tp}|$, $V_{t0} = 0.35$ V, $\gamma = 0.25\ V^{1/2}$, $2\phi_f = 1.0$ V, and $V_{DD} = 1$ V. Let Q_N and Q_P have dimensions $(W/L)_n = (W/L)_p = 130$ nm/65 nm. The total capacitance at the output node is 5 fF. Use as many of the results of Example 18.1 as you need.

(a) What are the values of V_{OH} and V_{OL}?

(b) For the situation in Fig. 18.7(a), find $i_{DN}(0)$, $i_{DP}(0)$, $i_{DN}(t_{PLH})$, $i_{DP}(t_{PLH})$, and t_{PLH}.

(c) For the situation depicted in Fig. 18.7(b), find $i_{DN}(0)$, $i_{DP}(0)$, $i_{DN}(t_{PHL})$, $i_{DP}(t_{PHL})$, and t_{PHL}. At what value of v_O will Q_P turn off?

(d) Find t_P.

Ans. (a) 1 V, 0 V; (b) 190.1 μA, 63.4 μA, 3.6 μA, 60 μA, 15.8 ps; (c) 190.1 μA, 63.4 μA, 180 μA, 1.2 μA, 11.5 ps, 0.41 V; (d) 13.7 ps

Equivalent Resistance of the Transmission Gate Although the transmission gate is capable of passing the full 1 and 0 levels to the load capacitance, it is not a perfect switch because it has a finite on-resistance. It is useful for us to obtain an estimate for this resistance. It can, for instance, be used together with the load capacitance as an alternative means to determining propagation delay. This approach is particularly useful in situations involving a network of inverters and transmission gates, as we shall shortly see.

To estimate the resistance of the transmission gate, we shall consider the situation in Fig. 18.7(a), where the transmission gate is on and is passing a high input (V_{DD}) to the capacitor load. Transistor Q_N operates in saturation until the output voltage v_O reaches $(V_{DD} - V_{tn})$, at which time Q_N turns off; thus,

$$i_{DN} = \frac{1}{2}k_n(V_{DD} - V_{tn} - v_O)^2 \quad \text{for} \quad v_O \le V_{DD} - V_{tn} \tag{18.8}$$

$$i_{DN} = 0 \quad \text{for} \quad v_O \ge V_{DD} - V_{tn} \tag{18.9}$$

A gross estimate for the equivalent resistance of Q_N can be obtained by dividing the voltage across it, $(V_{DD} - v_O)$, by i_{DN}, and neglecting the body effect, that is, assuming V_{tn} remains constant; thus,

$$R_{Neq} = \frac{V_{DD} - v_O}{\frac{1}{2}k_n(V_{DD} - V_{tn} - v_O)^2} \quad \text{for} \quad v_O \le V_{DD} - V_{tn} \tag{18.10}$$

and

$$R_{N\text{eq}} = \infty \qquad \text{for} \qquad v_O \geq V_{DD} - V_{tn} \tag{18.11}$$

Transistor Q_P will operate in saturation until $v_O = |V_{tp}|$, after which it enters the triode region; thus,

$$i_{DP} = \frac{1}{2}k_p\left(V_{DD} - |V_{tp}|\right)^2 \quad \text{for } v_O \leq |V_{tp}| \tag{18.12}$$

$$i_{DP} = k_p\left[\left(V_{DD} - |V_{tp}|\right)(V_{DD} - v_O) - \frac{1}{2}(V_{DD} - v_O)^2\right] \quad \text{for } v_O \geq |V_{tp}| \tag{18.13}$$

A gross estimate for the resistance of Q_P can be obtained by dividing the voltage across it, $(V_{DD} - v_O)$, by i_{DP}; thus,

$$R_{P\text{eq}} = \frac{V_{DD} - v_O}{\frac{1}{2}k_p\left(V_{DD} - |V_{tp}|\right)^2} \quad \text{for } v_O \leq |V_{tp}| \tag{18.14}$$

$$R_{P\text{eq}} = \frac{1}{k_p\left[V_{DD} - |V_{tp}| - \frac{1}{2}(V_{DD} - v_O)\right]} \quad \text{for } v_O \geq |V_{tp}| \tag{18.15}$$

Finally, the equivalent resistance R_{TG} of the transmission gate can be obtained as the parallel equivalent of $R_{N\text{eq}}$ and $R_{P\text{eq}}$,

$$R_{TG} = R_{N\text{eq}} \parallel R_{P\text{eq}} \tag{18.16}$$

Obviously, R_{TG} is a function of the output voltage v_O. As an example, we show in Fig. 18.8 a plot for R_{TG} for the transmission gate analyzed in Exercise 18.2, for the case of a high

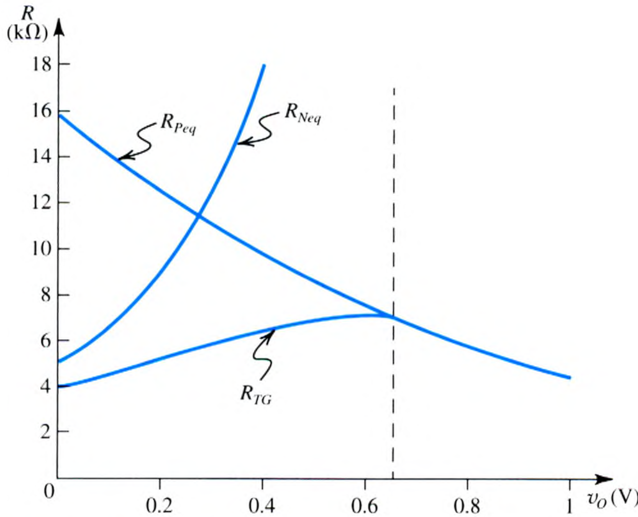


Figure 18.8 Plot of the equivalent resistances of the two transistors of the transmission gate in Fig. 18.7(a) passing a high input, and the overall resistance R_{TG} versus v_O. The data apply to the situation specified in Exercise 18.3.

input. Observe that R_{TG} remains relatively constant over the full range of v_O. The average value of R_{TG} over the range $v_O = 0$ to $V_{DD}/2$ can be used to determine t_{PLH}, as illustrated in Exercise 18.3.

EXERCISE

18.3 For the transmission gate analyzed in Exercise 18.2, whose equivalent resistance for capacitor charging is plotted in Fig. 18.8, use the average resistance value over the range $v_O = 0$ V to 0.5 V to determine t_{PLH}. Compare the result to the one we obtained using average currents in Exercise 18.2. Note that from the graph, $R_{TG} = 3.9$ kΩ at $v_O = 0$ V, and $R_{TG} = 7.1$ kΩ at $v_O = 0.5$ V. Recall that $t_{PLH} = 0.69RC$.

Ans. $t_{PLH} = 19.0$ ps, quite close to the value of 15.8 ps obtained in Exercise 18.2.

The expression for R_{TG} derived above applies only to the case of capacitor charging. A similar analysis can be performed for the case of capacitor discharge illustrated in Fig. 18.7(b). The resulting value of R_{TG} is close to the one we obtained earlier.

Having an estimate of the resistance of the transmission gate enables us to calculate the propagation delay of a signal path containing one or more transmission gates. Figure 18.9(a) shows one such circuit. It consists of a transmission gate connecting the output of an inverter to the input of another. We are interested in finding the propagation delay from the input of the first inverter to the input of the second as we apply a negative-going step to the input of the first inverter.

Fig. 18.9(b) shows the equivalent circuit where R_{P1} is the equivalent resistance of Q_{P1}, R_{TG} is the equivalent resistance of the transmission gate, C_{out1} is the output capacitance of the driver inverter, C_{TG1} and C_{TG2} are the capacitances introduced by the transmission gate at its input and output, respectively, and C_{in2} is the input capacitance of the load inverter. Observe that the circuit takes the form of an *RC* ladder network. A simple formula has been developed for calculating the delay of an arbitrarily long *RC* ladder network such as that shown in Fig. 18.10 having three sections. Known as the **Elmore delay formula**, it gives for the ladder in Fig. 18.10

$$t_P = 0.69[C_1R_1 + C_2(R_1 + R_2) + C_3(R_1 + R_2 + R_3)] \tag{18.17}$$

Applying the Elmore formula to the two-stage ladder in Fig. 18.9(b) gives

$$t_P = 0.69[(C_{\text{out1}} + C_{TG1})R_{P1} + (C_{\text{in2}} + C_{TG2})(R_{P1} + R_{TG})] \tag{18.18}$$

Finally, the factor 0.69 is usually dropped on the assumption that the input is a ramp rather than a step voltage.

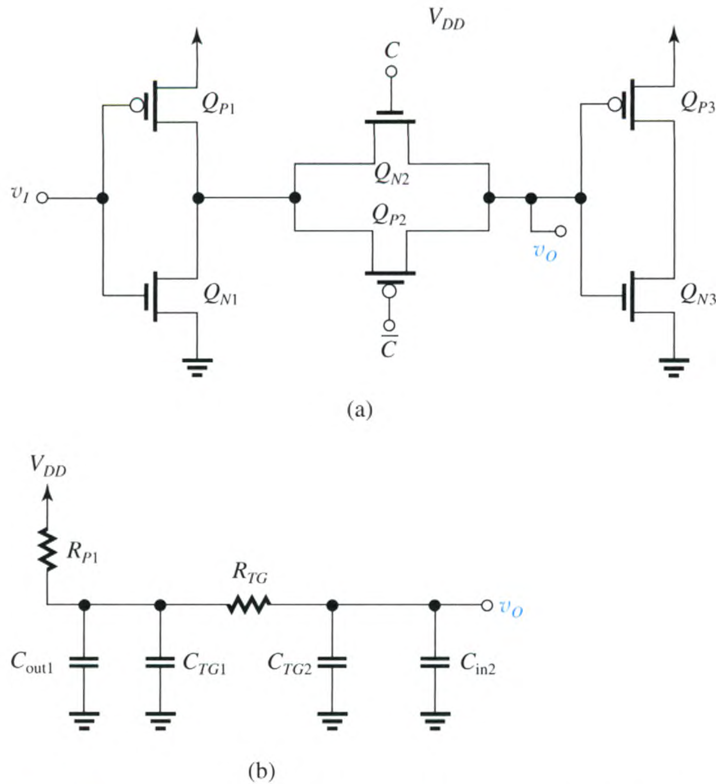


Figure 18.9 **(a)** A transmission gate connects the output of a CMOS inverter to the input of another. **(b)** Equivalent circuit for the purpose of analyzing the propagation delay of the circuit in **(a)**.

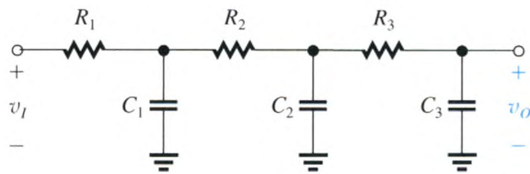


Figure 18.10 A three-section *RC* ladder network used to illustrate Elmore delay.

EXERCISE

18.4 The circuit in Fig. 18.9 is fabricated in a 65-nm CMOS technology; Q_{P1} of the first inverter has $W/L = 2$, and both transistors of the transmission gate have $W/L = 2$, as in Exercise 18.3. The capacitances are estimated to be $C_{out1} = 4$ fF, $C_{TG1} = C_{TG2} = 2$ fF, and $C_{in2} = 4$ fF. Estimate the average value of R_{P1} by calculating the current in Q_{P1} when its drain is at 0 V, and again when it is at 0.5 V. Also, use the average value of R_{TG} calculated in Exercise 18.3. Then, determine an estimate for t_{PLH} assuming a ramp input.

Ans. $R_{P1} = 12.1$ kΩ; $t_{PLH} = 0.18$ ns

18.2 Latches and Flip-Flops

In this section, we study the basic memory element, the latch, and consider some of its applications, including both static and dynamic circuits. As we will soon see, the pass transistors of Section 18.1 are often used to access internal storage nodes.

18.2.1 The Latch

The basic memory element, the latch, is shown in Fig. 18.11(a). It consists of two cross-coupled logic inverters, G_1 and G_2. The inverters form a positive-feedback loop. To investigate the operation of the latch we break the feedback loop at the input of one of the inverters, say G_1, and apply an input signal, v_W, as shown in Fig. 18.11(b). Assuming that the input impedance of G_1 is large, breaking the feedback loop will not change the loop voltage-transfer characteristic, which can be determined from the circuit of Fig. 18.11(b) by plotting v_Z versus v_W. This is the voltage-transfer characteristic of two cascaded inverters and thus takes the shape shown in Fig. 18.11(c). Observe that the transfer characteristic consists of three segments, with the middle segment corresponding to the transition region of the inverters.

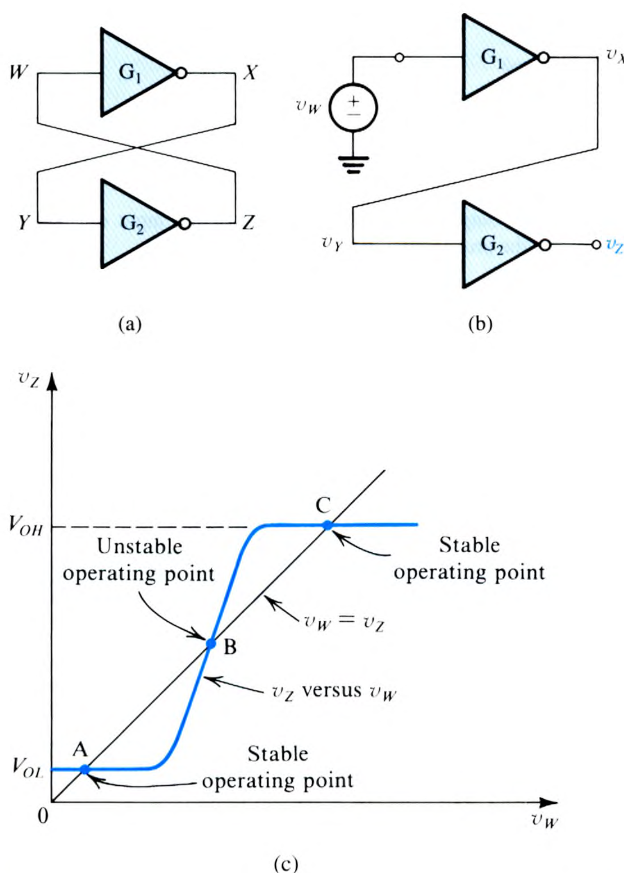


Figure 18.11 **(a)** Basic latch. **(b)** The latch with the feedback loop opened. **(c)** Determining the operating point(s) of the latch.

Also shown in Fig. 18.11(c) is a straight line with unity slope. This straight line represents the relationship $v_W = v_Z$ that is realized by reconnecting Z to W to close the feedback loop and thus to return it to its original form. As indicated, the straight line intersects the loop transfer curve at three points, A, B, and C. Thus any of these three points can serve as the operating point for the latch. We shall now show that while points A and C are stable operating points in the sense that the circuit can remain at either one indefinitely, point B is an unstable operating point; the latch cannot operate at B for any significant period of time.

We can see why point B is unstable if we consider the latch circuit in Fig. 18.11(a) to be operating at point B, and then take account of the electrical interference (or noise) that is inevitably present in any circuit. Let the voltage v_W increase by a small increment v_w. The voltage at X will increase (in magnitude) by a larger increment, equal to the product of v_w and the incremental gain of G_1 at point B. The resulting signal v_x is applied to G_2 and gives rise to an even larger signal at node Z. The voltage v_z is related to the original increment v_w by the loop gain at point B, which is the slope of the curve of v_Z versus v_W at point B. This gain is usually much greater than unity. Since v_z is coupled to the input of G_1, it becomes the new value of v_W and is further amplified by the loop gain. This regenerative process continues, shifting the operating point from B upward to point C, as illustrated in Fig. 18.12. Since at C the loop gain is zero (or almost zero), no regeneration can take place.

In the description above, we assumed arbitrarily an initial positive voltage increment at W. Had we instead assumed a negative voltage increment, we would have seen that the operating point moves downward from B to A. Again, since at point A the slope of the transfer curve is zero (or almost zero), no regeneration can take place. In fact, for regeneration to occur, the loop gain must be greater than unity, which is the case at point B. We can also model the positive feedback in the latch at operating point B as a negative feedback loop with $-180°$ phase shift and large gain at low frequencies, illustrated by the high slope of the loop voltage-transfer characteristic in Fig. 18.11(b). Thus, at operating point B the loop has no phase margin and is unstable.

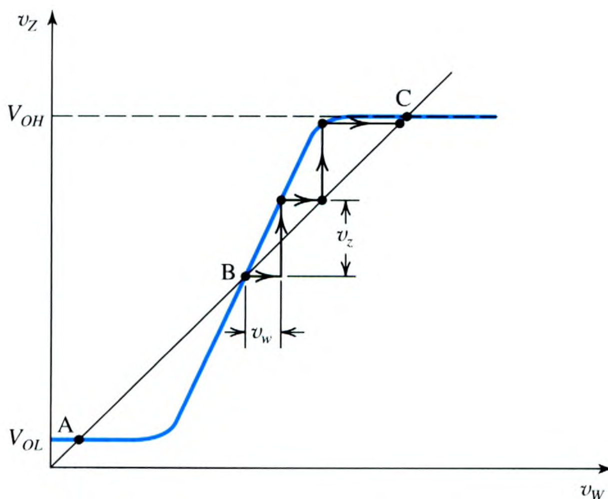


Figure 18.12 Point B is an unstable operating point for the latch: A small positive increment v_w gets amplified around the loop and causes the operating point to shift to the stable operating point C. Had v_w been negative, the operating point would have shifted to the other stable point, A.

The discussion above leads us to conclude that the latch has two stable operating points, A and C. At point C, v_W is high, v_X is low, v_Y is low, and v_Z is high. The reverse is true at point A. If we consider X and Z as the latch outputs, we see that in one of the stable states (say that corresponding to operating point A), v_X is high (at V_{OH}) and v_Z is low (at V_{OL}). In the other state (corresponding to operating point C), v_X is low (at V_{OL}) and v_Z is high (at V_{OH}). Thus the latch is a **bistable** circuit having two complementary outputs. The stable state in which the latch operates depends on the external excitation that forces it to the particular state. The latch then *memorizes* this external action by staying indefinitely in the acquired state. As a memory element the latch is capable of storing one bit of information. For instance, we can arbitrarily designate the state in which v_X is high and v_Z is low as corresponding to a stored logic 1. The other complementary state then is designated by a stored logic 0. Finally, we note that the latch circuit described is of the static variety.

We still need to devise a mechanism by which the latch can be *triggered* to change state. The latch together with the triggering circuitry forms a *flip-flop*, which we discuss next. We also present analog bistable circuits using op amps in Chapter 15.

18.2.2 The SR Flip-Flop

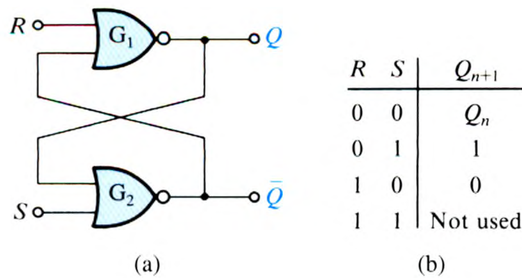
The simplest type of flip-flop is the set/reset (SR) flip-flop shown in Fig. 18.13(a). It is formed by cross-coupling two NOR gates, and thus it incorporates a latch. The second inputs of G_1 and G_2 together serve as the trigger inputs of the flip-flop. These two inputs are labeled S (for set) and R (for reset). The outputs are labeled Q and $\overline{Q}$, emphasizing their complementarity. The flip-flop is considered to be set (i.e., storing a logic 1) when Q is high and $\overline{Q}$ is low. When the flip-flop is in the other state (Q low, $\overline{Q}$ high), it is considered to be reset (storing a logic 0).

In the *rest* or *memory state* (i.e., when we do not wish to change the state of the flip-flop), both the S and R inputs should be low. Consider the case when the flip-flop is storing a logic 0. Since Q will be low, both inputs to the NOR gate G_2 will be low. Its output will therefore be high. This high is applied to the input of G_1, causing its output Q to be low, satisfying the original assumption. To set the flip-flop we raise S to the logic-1 level while leaving R at 0.

FLIP-FLOP FACT

In 1918 William H. Eccles and Frank W. Jordan, while working at City and Guilds Technical College in London, filed the first patent for an electronic bistable element, the bellwether of the digital electronic age. The initial design captured the idea of a closed positive-feedback loop involving two inverting direct-coupled amplifiers, along with “gating” to allow signals to be injected. While the first designs used vacuum tubes, the same structure was extended to bipolar and MOS transistors as those technologies became commercially available. The Eccles–Jordan flip-flop remains, over a century later, an enormous milestone in digital computing.

The 1 at the S terminal will force the output of G_2, $\overline{Q}$, to 0. Thus the two inputs to G_1 will be 0 and its output Q will go to 1. Now even if S returns to 0, the $Q = 1$ signal fed to the input of G_2 will keep $\overline{Q} = 0$, and the flip-flop will remain in the newly acquired set state. Note that if we raise S to 1 again (with R remaining at 0), no change will occur. To reset the flip-flop we need to raise R to 1 while leaving $S = 0$. We can readily show that this forces the flip-flop into the reset state ($Q = 0, \overline{Q} = 1$) and that the flip-flop remains in this state even



R	S	Q_{n+1}
0	0	Q_n
0	1	1
1	0	0
1	1	Not used

(b)

Figure 18.13 **(a)** The set/reset (SR) flip-flop and **(b)** its truth table.

after R has returned to 0. Notice that the trigger signal merely starts the regenerative action of the positive-feedback loop of the latch.

Finally, let's see what happens if both S and R are simultaneously raised to 1. The two NOR gates will cause both Q and $\overline{Q}$ to become 0 (note that in this case the complementary labeling of these two variables is incorrect). However, if R and S return to the rest state ($R = S = 0$) simultaneously, the state of the flip-flop will be undefined. In other words, it will be impossible to predict the final state of the flip-flop. For this reason, this input combination is usually disallowed (i.e., not used). Note, however, that this situation arises only in the idealized case, when both R and S return to 0 precisely simultaneously. In actual practice one of the two will return to 0 first, and the final state will be determined by the input that remains high longest.

The operation of the flip-flop is summarized by the *truth table* in Fig. 18.13(b), where Q_n denotes the value of Q at time t_n just before the application of the R and S signals, and Q_{n+1} denotes the value of Q at time t_{n+1} after the application of the input signals.

Rather than using two NOR gates, we can also implement an SR flip-flop by cross-coupling two NAND gates, in which case the set and reset functions are active when low (see Problem 18.16).

18.2.3 CMOS Implementation of SR Flip-Flops

The SR flip-flop of Fig. 18.13 can be directly implemented in CMOS by simply replacing each of the NOR gates with its CMOS circuit realization. We encourage you to sketch the resulting circuit (see Problem 18.15). Although the CMOS circuit thus obtained works well, it is somewhat complex. As an alternative, we consider a simplified circuit that furthermore implements additional logic. Specifically, Fig. 18.14 shows a *clocked* version of an SR flip-flop. Since the clock inputs form AND functions with the set and reset inputs, the flip-flop can be set or reset only when the clock ϕ is high. Observe that although the two cross-coupled inverters at the heart of the flip-flop are of the standard CMOS type, only NMOS transistors are used for the set–reset circuitry. Nevertheless, since there is no conducting path between V_{DD} and ground (except during switching), the circuit does not dissipate any static power.

Except for the addition of clocking, the SR flip-flop of Fig. 18.14 operates in exactly the same way as its logic antecedent in Fig. 18.13: To illustrate, consider what happens when the flip-flop is in the reset state ($Q = 0, \overline{Q} = 1$, $v_Q = 0$, $v_{\overline{Q}} = V_{DD}$), and assume that we wish to set it. To do so, we arrange for a high (V_{DD}) signal to appear on the S input while R is held low at 0 V. Then, when the clock ϕ goes high, both Q_5 and Q_6 will conduct, pulling the voltage $v_{\overline{Q}}$ down. If $v_{\overline{Q}}$ goes below the threshold V_M of the (Q_3, Q_4) inverter, the inverter will switch states (or at least begin to switch states), and its output v_Q will rise. This increase in v_Q is fed

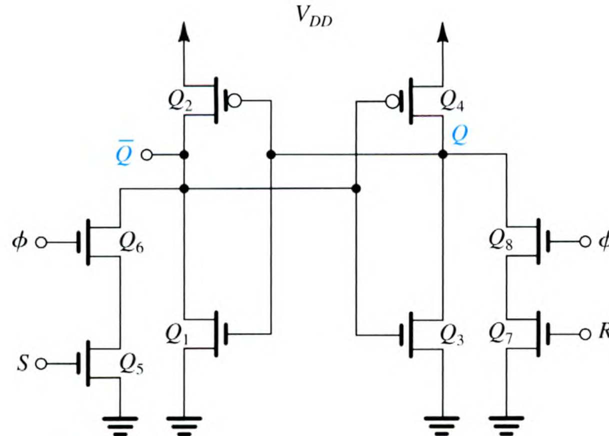


Figure 18.14 CMOS implementation of a clocked SR flip-flop. The clock signal is denoted by ϕ.

back to the input of the (Q_1, Q_2) inverter, causing its output $v_{\overline{Q}}$ to go down even further; the regeneration process, characteristic of the positive-feedback latch, is now in progress.

The preceding description of flip-flop switching is predicated on two assumptions:

1. Transistors Q_5 and Q_6 supply sufficient current to pull the node $\overline{Q}$ down to a voltage at least slightly below the threshold of the (Q_3, Q_4) inverter. This is essential for the regenerative process to begin. Without this initial trigger, the flip-flop will fail to switch. In Example 18.2, we shall investigate the *minimum W/L* ratios that Q_5 and Q_6 must have to meet this requirement.
2. The set signal remains high long enough to cause regeneration to take over the switching process. We can estimate the minimum width required for the set pulse by adding the interval during which $v_{\overline{Q}}$ is reduced from V_{DD} to $V_{DD}/2$ and the interval for the voltage v_Q to respond and rise to $V_{DD}/2$. This point also will be illustrated in Example 18.2.

Finally, note that the symmetry of the circuit indicates that all the preceding remarks apply equally well to the reset process.

Example 18.2

The CMOS SR flip-flop in Fig. 18.14 is fabricated in a 65-nm process for which $\mu_n C_{ox} = 3\mu_p C_{ox} = 450\ \mu\text{A/V}^2$, $V_{tn} = -V_{tp} = 0.35$ V, and $V_{DD} = 1.0$ V. The inverters have $(W/L)_n = 130$ nm/65 nm and $(W/L)_p = 3(W/L)_n$. The four NMOS transistors in the set–reset circuit, Q_{5-8}, have equal W/L ratios.

(a) Determine the minimum value required for this ratio to ensure that the flip-flop will switch.
(b) Determine the minimum width the set pulse must have for the case in which the W/L ratio of each of the four transistors in the set–reset circuit is selected at twice the minimum value found in (a). Assume that the total capacitance between each of the Q and $\overline{Q}$ nodes and ground is 10 fF.

Solution

(a) Figure 18.15(a) shows the portion of the circuit we need for our purposes. Notice that since the circuit is in the reset state and regeneration has not yet begun, we can assume that $v_Q = 0$ and thus Q_2 is conducting. The circuit is in effect a pseudo-NMOS gate (introduced in Section 16.2), and our task is to select the W/L ratios for Q_5 and Q_6 so that V_{OL} of this inverter is lower than $V_{DD}/2$ (the threshold of the Q_3, Q_4 inverter whose Q_N and Q_P are matched). We can find the minimum required W/L for Q_5 and Q_6 by equating the current supplied by Q_5 and Q_6 to the current supplied by Q_2 at $v_{\bar{Q}} = V_{DD}/2$. To simplify matters, we assume that the series connection of Q_5 and Q_6 is equivalent to a single transistor whose W/L is half the W/L of each of Q_5 and Q_6 [Fig. 18.15(b)]. Now, since at $v_{\bar{Q}} = V_{DD}/2 = 0.5$ V and $|V_t| = 0.35$ V, both this equivalent transistor and Q_2 operate in the triode region, we can write

$$I_{Deq} = I_{D2}$$

$$450 \times \frac{1}{2}\left(\frac{W}{L}\right)_5 \left[(1-0.35)\left(\frac{1.0}{2}\right) - \frac{1}{2}\left(\frac{1.0}{2}\right)^2\right] = 150 \times \frac{390}{65}\left[(1-0.35)\left(\frac{1.0}{2}\right) - \frac{1}{2}\left(\frac{1.0}{2}\right)^2\right]$$

which yields

$$\left(\frac{W}{L}\right)_5 = 260 \text{ nm}/65 \text{ nm}$$

and thus

$$\left(\frac{W}{L}\right)_6 = 260 \text{ nm}/65 \text{ nm}$$

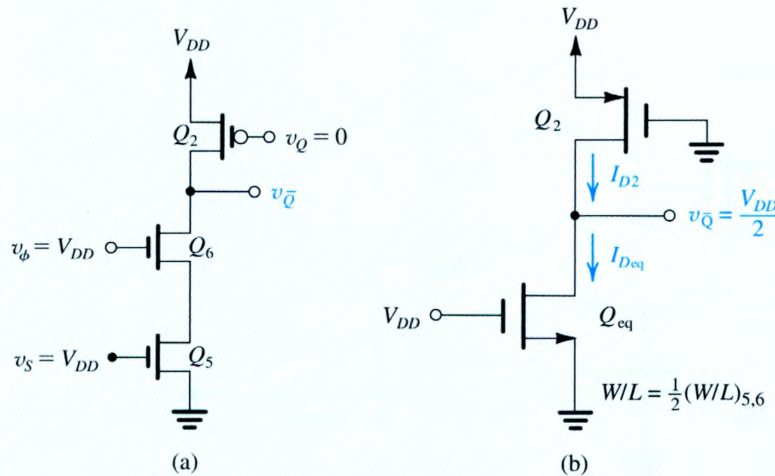


Figure 18.15 **(a)** The relevant portion of the flip-flop circuit of Fig. 18.14 for determining the minimum W/L ratios of Q_5 and Q_6 needed to ensure that the flip-flop will switch. **(b)** The circuit in (a) with Q_5 and Q_6 replaced with their equivalent transistor Q_{eq}, at the point of switching.

Example 18.2 *continued*

(b) The value calculated for $(W/L)_5$ and $(W/L)_6$ is the absolute minimum needed for switching to occur. To guarantee that the flip-flop will switch, the value selected for $(W/L)_5$ and $(W/L)_6$ is usually somewhat larger than the minimum. Selecting a value twice the minimum,

$$(W/L)_5 = (W/L)_6 = 520 \text{ nm}/65 \text{ nm}$$

The minimum required width of the set pulse is composed of two components: the time for $v_{\bar{Q}}$ in Fig. 18.15(a) to fall from V_{DD} to $V_{DD}/2$, where $V_{DD}/2$ is the threshold voltage of the inverter formed by Q_3 and Q_4 in Fig. 18.14; and the time for the output of the Q_3-Q_4 inverter to rise from 0 to $V_{DD}/2$. At the end of the second interval, the feedback signal will have traveled around the feedback loop, and regeneration can continue without the presence of the set pulse. We denote the first component t_{PHL} and the second t_{PLH}, and calculate their values as follows.

To determine t_{PHL}, refer to the circuit in Fig. 18.16 and note that the capacitor discharge current i_C is the difference between the current of the equivalent transistor Q_{eq} and the current of Q_2,

$$i_C = i_{Deq} - i_{D2}$$

To determine the average discharge current i_C, calculate i_{Deq} and i_{D2} at $t = 0$ and $t = t_{PHL}$. At $t = 0$, $v_{\bar{Q}} = V_{DD}$, thus Q_2 is off,

$$i_{D2}(0) = 0$$

and Q_{eq} is in saturation,

$$i_{Deq} = \frac{1}{2} \times 450 \times \frac{1}{2} \times \frac{520}{65} \times (1.0 - 0.35)^2 = 380.3 \text{ μA}$$

Thus,

$$i_C(0) = 380.3 - 0 = 380.3 \text{ μA}$$

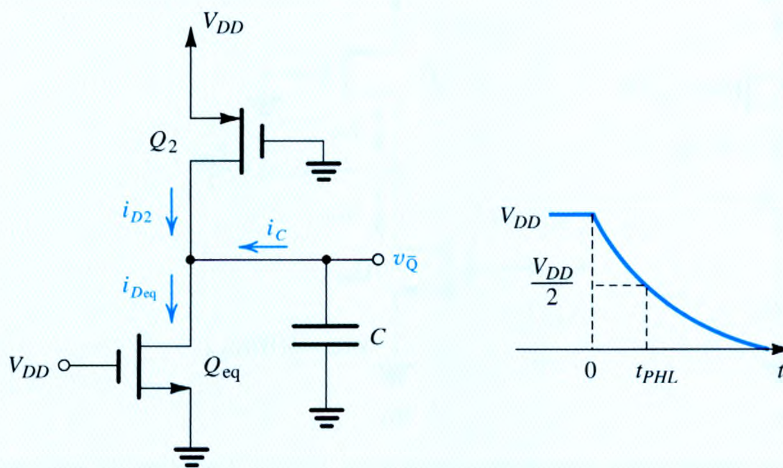


Figure 18.16 Determining the time t_{PHL} for $v_{\bar{Q}}$ to fall from V_{DD} to $V_{DD}/2$.

At $t = t_{PHL}$, $v_{\overline{Q}} = V_{DD}/2$, thus both Q_2 and Q_{eq} are in the triode region,

$$i_{D2}(t_{PHL}) = 150 \times \frac{390}{65} \times \left[(1-0.35)\left(\frac{1.0}{2}\right) - \frac{1}{2}\left(\frac{1.0}{2}\right)^2\right] = 180\ \mu\text{A}$$

and

$$i_{D_{eq}}(t_{PHL}) = 450 \times \frac{1}{2} \times \frac{520}{65} \times \left[(1-0.35)\left(\frac{1.0}{2}\right) - \frac{1}{2}\left(\frac{1.0}{2}\right)^2\right] = 360\ \mu\text{A}$$

Thus,

$$i_C(t_{PHL}) = 360 - 180 = 180\ \mu\text{A}$$

and the average value of i_C over the interval $t = 0$ to $t = t_{PHL}$ is

$$i_C|_{av} = \frac{i_C(0) + i_C(t_{PHL})}{2} = \frac{380.3 + 180}{2} = 280.1\ \mu\text{A}$$

We now can calculate t_{PHL} as

$$t_{PHL} = \frac{C(V_{DD}/2)}{i_C|_{av}} = \frac{10 \times 10^{-15} \times 0.5}{280.1 \times 10^{-6}} = 17.9\ \text{ps}$$

Next we consider the time t_{PLH} for the output of the Q_3-Q_4 inverter, v_Q, to rise from 0 to $V_{DD}/2$. We can calculate the value of t_{PLH} using the propagation-delay formula derived in Chapter 17 (Eq. 17.10), namely,

$$t_{PLH} = \frac{\alpha_p C}{k_p'(W/L)_p V_{DD}}$$

where

$$\alpha_p = \frac{2}{\frac{7}{4} - \frac{3|V_{tp}|}{V_{DD}} + \left(\frac{|V_{tp}|}{V_{DD}}\right)^2}$$

Substituting numerical values gives us

$$\alpha_p = \frac{2}{1.75 - \frac{3\times 0.35}{1.0} + \left(\frac{0.35}{1.0}\right)^2} = 2.43$$

and

$$t_{PLH} = \frac{2.43 \times 10 \times 10^{-15}}{150 \times 10^{-6} \times (390/65) \times 1.0} = 27.0\ \text{ps}$$

Finally, we can calculate the minimum required width of the set pulse:

$$T_{min} = t_{PHL} + t_{PLH} = 44.9\ \text{ps}$$

EXERCISE

18.5 For the SR flip-flop specified in Example 18.2, find the minimum W/L for both Q_5 and Q_6 so that switching is achieved when inputs S and ϕ are at $(V_{DD}/2)$.
Ans. 71.1

18.2.4 A Simpler CMOS Implementation of the Clocked SR Flip-Flop

A simpler implementation of a clocked SR flip-flop is shown in Fig. 18.17. Here, pass-transistor logic is employed to implement the clocked set–reset functions. This circuit is very popular in the design of static random-access memory (SRAM) chips, where it is used as the basic memory cell (Section 18.3).

18.2.5 D Flip-Flop Circuits

A variety of flip-flop types exist and can be synthesized using logic gates. We can obtain CMOS circuit implementations by simply replacing the gates with their CMOS circuit realizations. This approach, however, usually results in rather complex circuits. In many cases, we can find simpler circuits by taking a circuit-design viewpoint, rather than a logic-design one. To illustrate this point, let's consider the CMOS implementation of a very important type of flip-flop, the data, or D, flip-flop.

The D flip-flop is shown in block diagram form in Fig. 18.18. It has two inputs, the data input D and the clock input ϕ. The complementary outputs are labeled Q and $\overline{Q}$. When the clock is low, the flip-flop is in the memory, or rest, state; signal changes on the D input line have no effect on the state of the flip-flop. As the clock goes high, the flip-flop acquires the logic level that existed on the D line just before the rising edge of the clock. Such a flip-flop

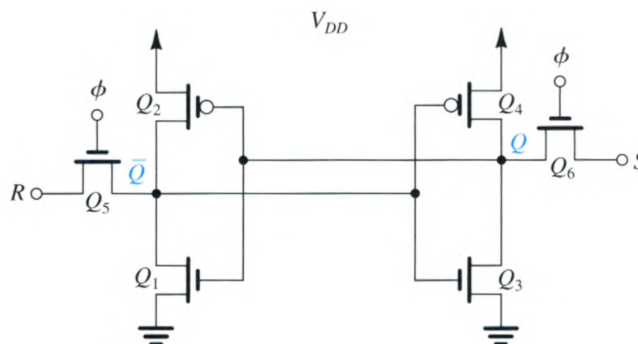


Figure 18.17 A simpler CMOS implementation of the clocked SR flip-flop. This circuit is popular as the basic cell in the design of static random-access memory (SRAM) chips.

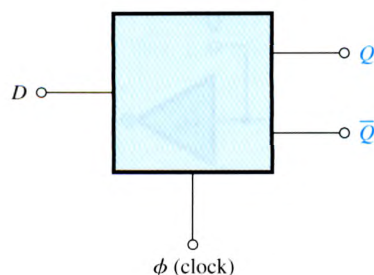


Figure 18.18 A block diagram representation of the D flip-flop.

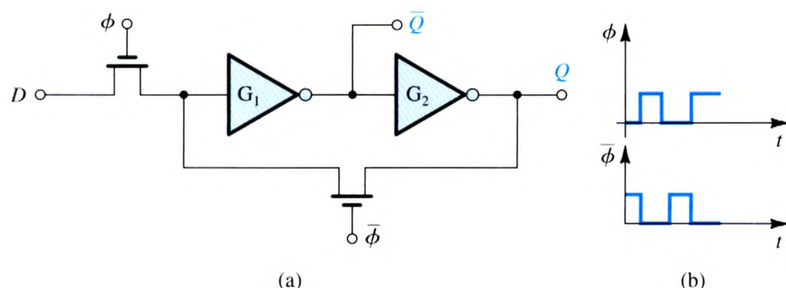


Figure 18.19 A simple implementation of the D flip-flop. The circuit in **(a)** utilizes the two-phase **nonoverlapping clock** whose waveforms are shown in **(b).**

is said to be **edge triggered**. Some implementations of the D flip-flop include direct set and reset inputs that override the clocked operation just described.

A simple implementation of the D flip-flop is shown in Fig. 18.19. The circuit consists of two inverters connected in a positive-feedback loop, just as in the static latch of Fig. 18.11(a). Here, however, the loop is closed for only part of the time—specifically, when the clock is low ($\phi = 0, \overline{\phi} = 1$). The input D is connected to the flip-flop through a switch that closes when the clock is high. Operation is straightforward: When ϕ is high, the loop is opened, and the input D is connected to the input of inverter G_1. The capacitance at the input node of G_1 is charged to the value of D, and the capacitance at the input node of G_2 is charged to the value of $\overline{D}$. Then, when the clock goes low, the input line is isolated from the flip-flop, the feedback loop is closed, and the latch acquires the state corresponding to the value of D just before ϕ went down, providing an output $Q = D$.

From this we can see that the circuit in Fig. 18.19 combines the positive-feedback technique of static bistable circuits and a charge-storage technique found in dynamic circuits. It is important to note that the proper operation of this circuit, and of many circuits that use clocks, is predicated on the assumption that *ϕ and $\overline{\phi}$ will not be simultaneously high at any time*. This condition is defined by referring to the two clock phases as being *nonoverlapping*.

An inherent drawback of the D flip-flop implementation of Fig. 18.19 is that during ϕ, the output of the flip-flop simply follows the signal on the D input line. This can cause problems in certain logic-design situations. The problem is solved very effectively by using the **master–slave** configuration shown in Fig. 18.20(a). Before discussing how the circuit operates, we note that although the switches are shown implemented with single NMOS transistors, CMOS transmission gates are employed in many applications. We are simply using the single MOS transistor as a "shorthand notation" for a series switch.

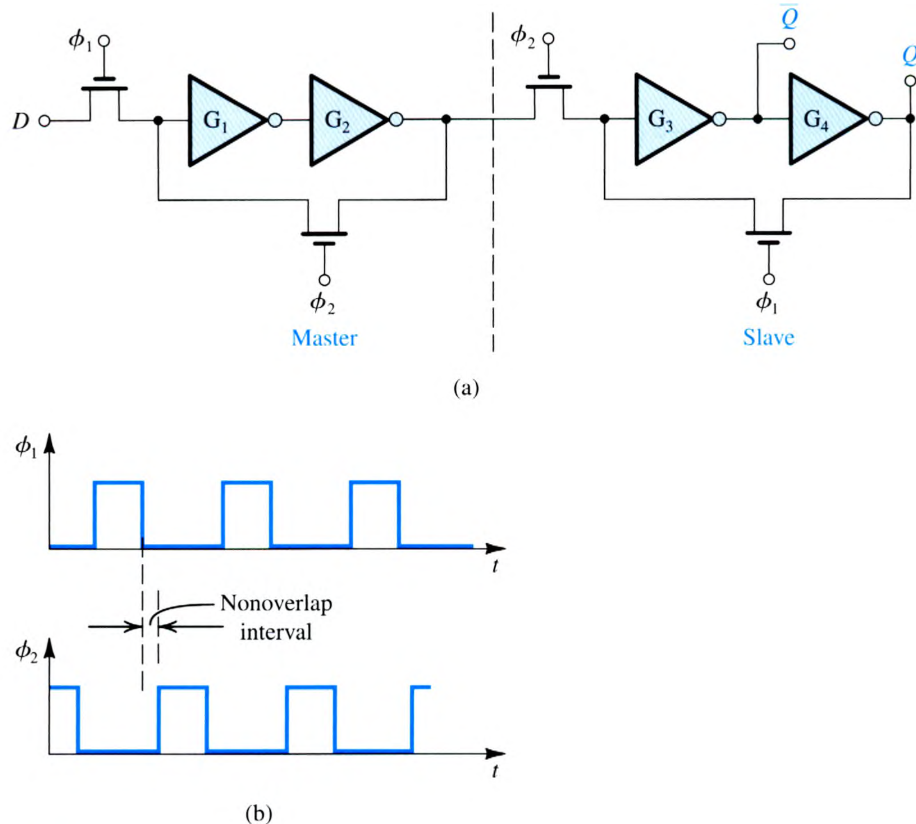


Figure 18.20 **(a)** A master–slave D flip-flop. The switches can be, and usually are, implemented with CMOS transmission gates. **(b)** Waveforms of the two-phase nonoverlapping clock required.

The master–slave circuit consists of a pair of circuits of the type shown in Fig. 18.19, operated with alternate clock phases. Here, to emphasize that the two clock phases must be nonoverlapping, we denote them ϕ_1 and ϕ_2, and clearly show the nonoverlap interval in the waveforms of Fig. 18.20(b). This is how the circuit operates:

1. When ϕ_1 is high and ϕ_2 is low, the input is connected to the master latch, whose feedback loop is opened, while the slave latch is isolated. Thus, the output Q remains at the value stored previously in the slave latch, whose loop is now closed. The node capacitances of the master latch are charged to the appropriate voltages corresponding to the present value of D.
2. When ϕ_1 goes low, the master latch is isolated from the input data line. Then, when ϕ_2 goes high, the feedback loop of the master latch is closed, locking in the value of D. Further, its output is connected to the slave latch, whose feedback loop is now open. The node capacitances in the slave are appropriately charged so that when ϕ_1 goes high again, the slave latch locks in the new value of D and provides it at the output, $Q = D$.

From this description, we note that at the positive transition of clock ϕ_2 the output Q adopts the value of D that existed on the D line at the end of the preceding clock phase,

ϕ_1. This output value remains constant for one clock period. Finally, note that during the nonoverlap interval both latches have their feedback loops open, and we are relying on the node capacitances to maintain most of their charge. It follows that the nonoverlap interval should be kept reasonably short (perhaps equivalent to the delay of one or two logic gates, such that the held state will not decay due to leakage).

18.3 Random-Access Memory (RAM) Cells

A computer system, whether a large machine in a data center, a laptop, a smartphone, or an embedded device, requires memory for storing data and program instructions. Broadly speaking, memory can be divided into two types: main memory and mass-storage memory. The main memory is usually the most rapidly accessible memory and the one from which most instructions in programs are executed. A **random-access memory** (RAM) is one in which the time required for storing (writing) information and for retrieving (reading) information is independent of the physical location (within the memory) in which the information is stored. Main memory is usually made up of RAM. RAMs should be contrasted with *serial* or *sequential* memories, such as disks and tapes, from which data are available only in the sequence in which they were originally stored.

Another important classification of memory relates to whether it is a read/write or a read-only memory. **Read/write** (R/W) memory permits data to be stored and retrieved at comparable speeds. Computer systems require random-access read/write memory for data and program storage. **Read-only memories** (ROM) permit reading at the same high speeds as R/W memories (or perhaps higher) but restrict the writing operation. ROMs can be used to store a microprocessor operating-system program. They are also employed in operations that require table lookup, such as finding the values of mathematical functions.

A third classification depends on whether data that are stored will be destroyed should the power supply to the memory be cut off. **Volatile** memory, such as **static RAM** (SRAM) and **dynamic RAM** (DRAM), must be powered up to hold their data. On the other hand, **nonvolatile** memory such as **Flash** memory (typically organized into a RAM-type array) relies on circuit techniques that keep data safely stored when devices are turned off.

The regular structure of RAM-type memory arrays makes them ideal for dense storage of information. Indeed, at any moment, memory chips represent the state of the art in integrated circuit packing density. Beginning with the introduction of a 1-Kbit chip in 1970, memory-chip density has quadrupled about every 3 years. At the present time (2020), 16-Gbit DRAM chips are readily available. In the past decade, Flash memory—also based on dense packing of storage cells onto an integrated circuit—has largely displaced disks and tapes for mass storage. In this section, we study some of the basic circuits used in SRAMs, DRAMs, and Flash memory.

The bulk of a memory chip consists of cells in which bits are stored. Each memory cell is an electronic circuit capable of storing one bit (or possibly a few). For many reasons, it is desirable to have the storage cells on a chip arranged in a square or a nearly square matrix. Figure 18.21 illustrates this arrangement. The cell matrix has 2^M rows and 2^N columns, for a total storage capacity of 2^{M+N}. For example, a 1-Mbit square matrix would have 1024 rows and 1024 columns ($M = N = 10$). Each cell in the array is connected to one of the 2^M row lines, known rather loosely, but universally, as **word lines**, and to one of the 2^N column lines, known as **bit lines**. A particular cell is selected for reading or writing by activating its word

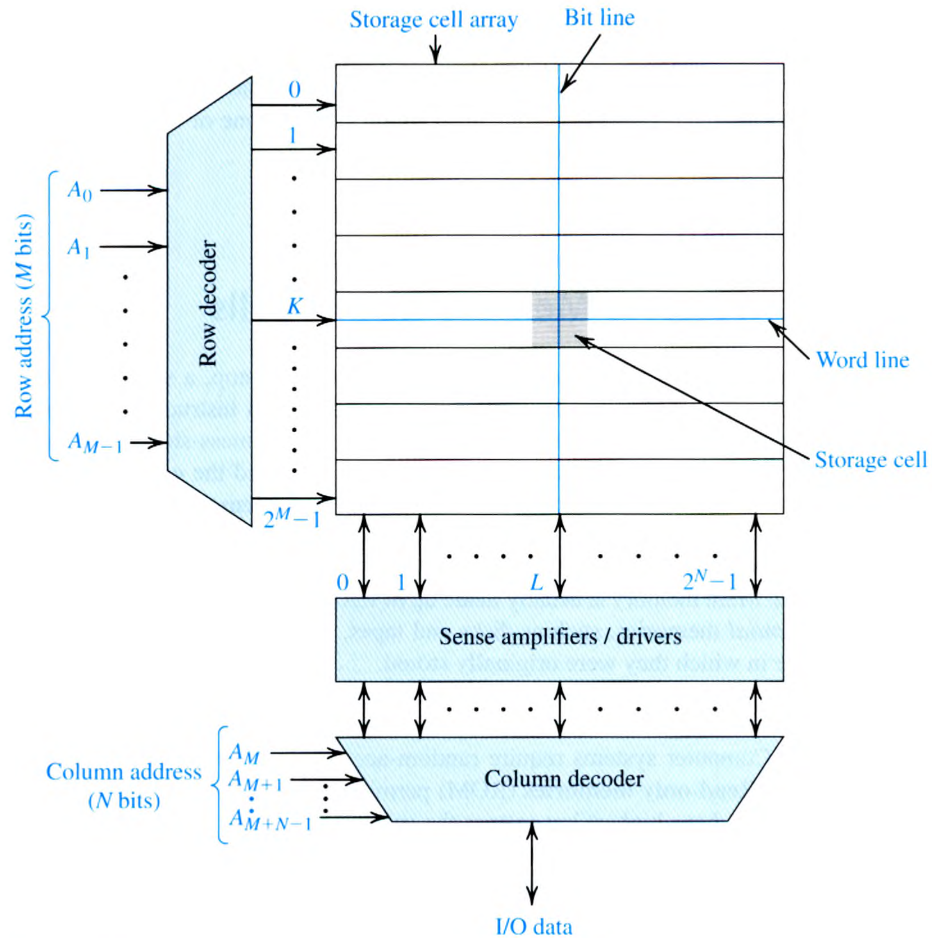


Figure 18.21 A 2^{M+N}-bit memory chip organized as an array of 2^M rows × 2^N columns.

line and its bit line. How this is done depends on whether the memory is static or dynamic, as we will see shortly.

A 2^M word line is activated by the **row decoder**, a combinational logic circuit that selects (raises the voltage of) the particular word line whose M-bit address is applied to the decoder input. When the Kth word line is activated for, say, a read operation, all 2^N cells in row K provide their contents to their respective bit lines. Thus, if the cell in column L (Fig. 18.21) is storing a 1, the voltage of bit-line number L will be raised, usually by a small voltage, say 0.1 to 0.2 V. The readout voltage is small because the cell is small, a deliberate design decision, since the number of cells is very large. The small readout signal is applied to a **sense amplifier** connected to the bit line, which provides a full-swing digital signal (from 0 to V_{DD}) at its output. This signal, together with the output signals from all the other cells in the selected row, is then delivered to a column decoder, which selects the signal of the particular column whose N-bit address is applied to the decoder input and causes this signal to appear on the chip input/output (I/O) data line.

Since the major part of the memory chip is taken up by the storage cells, it follows that to be able to pack a large number of bits on a chip, it is imperative that the cell be made as

small as possible. The power dissipation per cell should also be minimized. Thus, many of the flip-flop circuits studied in Section 18.2 are too complex to be suitable for implementing the storage cells in a RAM chip. However, as we will see soon, SRAMs use static latches as the storage cells. On the other hand, DRAMs store their data on capacitors, resulting in further reduction in cell area, but at the expense of more complex read and write circuitry.

In particular, while SRAMs can hold their stored data indefinitely, provided the power supply remains on, DRAMs require *periodic refreshing* to regenerate the data stored on capacitors. This is because the storage capacitors discharge, though slowly, as a result of leakage currents. By virtue of their smaller cell size, dynamic memory chips are typically four times as dense as their contemporary static chips. In the following subsections, we will study basic SRAM and DRAM storage cells.

18.3.1 Static Memory (SRAM) Cell

Figure 18.22 shows a typical static memory cell in CMOS technology. The circuit, which we encountered in Section 18.2, is a flip-flop comprising two cross-coupled inverters and two **access transistors**, Q_5 and Q_6. The access transistors are turned on when the word line is selected and its voltage raised to V_{DD}, and they connect the flip-flop to the column (bit or B) line and column (bit or $\overline{B}$) line. Note that although in principle only the B or the $\overline{B}$ line suffices, most often both are utilized, as shown in Fig. 18.22. This practice provides a *differential data path* between the cell and the memory-chip output while increasing the circuit reliability. The access transistors act as transmission gates allowing bidirectional current flow between the flip-flop and the B and $\overline{B}$ lines. This circuit is known as the **six-transistor** or **6T SRAM cell**.

The Read Operation Consider first a read operation, and assume that the cell is storing a 1. In this case, Q will be high at V_{DD}, and $\overline{Q}$ will be low at 0 V. Before the read operation begins, the B and $\overline{B}$ lines are raised to a voltage in the range $V_{DD}/2$ to V_{DD}. This process, known as **precharging**, is performed using circuits we will discuss in Section 18.4, where we consider sense amplifiers. To simplify matters, let's assume here that the precharge voltage of B and $\overline{B}$ is V_{DD}.

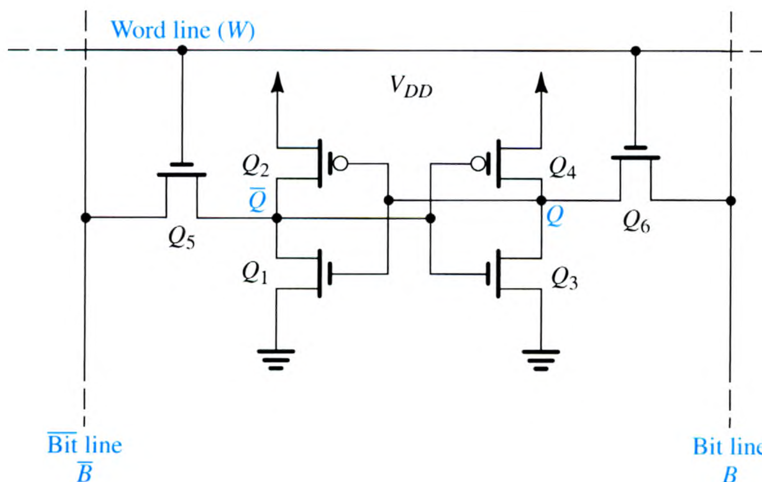


Figure 18.22 A CMOS SRAM memory cell.

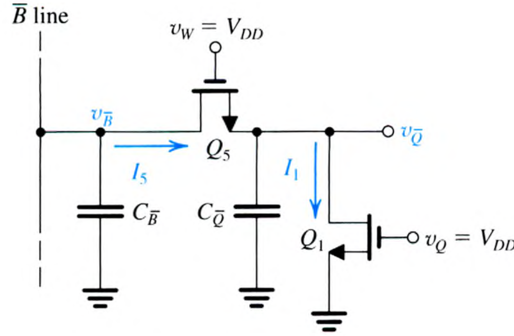


Figure 18.23 Relevant parts of the SRAM cell circuit during a read operation when the cell is storing a logic 1. Note that initially $v_Q = V_{DD}$ and $v_{\overline{Q}} = 0$. Also note that the B and $\overline{B}$ lines are precharged to a voltage V_{DD}.

When the word line is selected and the access transistors Q_5 and Q_6 are turned on, we can see that the only portion of the circuit that will be conducting is what is shown in Fig. 18.23. Noting that the initial value of $v_{\overline{Q}}$ is 0 V, we can see that current will flow from the $\overline{B}$ line (actually, from the $\overline{B}$-line capacitance $C_{\overline{B}}$) through Q_5 and into capacitor $C_{\overline{Q}}$, which is the small equivalent capacitance between the $\overline{Q}$ node and ground. This current charges $C_{\overline{Q}}$ and thus $v_{\overline{Q}}$ rises and Q_1 conducts, sinking some of the current supplied by Q_5. Equilibrium is reached when $C_{\overline{Q}}$ is charged to a voltage $V_{\overline{Q}}$ at which I_1 equals I_5, and no current flows through $C_{\overline{Q}}$. Here it is extremely important to note that to avoid changing the state of the flip-flop, that is, for our read operation to be **nondestructive**, $V_{\overline{Q}}$ must not exceed the threshold voltage of the inverter Q_3–Q_4. In fact, SRAM designers usually impose a more stringent requirement on the value of $V_{\overline{Q}}$, namely, that it should be lower than the threshold voltage of Q_3, V_{tn}. Thus, the design problem we shall now solve is as follows: Determine the ratio of $(W/L)_5/(W/L)_1$ so that $V_{\overline{Q}} \le V_{tn}$.

Noting that Q_5 will be operating in saturation and neglecting, for simplicity, the body effect, we can write

$$I_5 = \frac{1}{2}(\mu_n C_{ox})\left(\frac{W}{L}\right)_5 (V_{DD} - V_{tn} - V_{\overline{Q}})^2 \tag{18.19}$$

Transistor Q_1 will be operating in the triode region, and its current I_1 can be written as

$$I_1 = (\mu_n C_{ox})\left(\frac{W}{L}\right)_1 \left[(V_{DD} - V_{tn})V_{\overline{Q}} - \frac{1}{2}V_{\overline{Q}}^2\right] \tag{18.20}$$

Equating I_5 and I_1 gives a quadratic equation in $V_{\overline{Q}}$, which can be solved to obtain

$$\frac{V_{\overline{Q}}}{V_{DD} - V_{tn}} = 1 - 1\Bigg/\sqrt{1 + \frac{(W/L)_5}{(W/L)_1}} \tag{18.21}$$

This is an attractive relationship, since it provides $V_{\overline{Q}}$ in normalized form and thus always applies, independent of the process technology utilized. Figure 18.24 shows a universal plot of $\left[V_{\overline{Q}}/(V_{DD} - V_{tn})\right]$ versus $(W/L)_5/(W/L)_1$. For a given process technology, V_{DD} and V_{tn} are determined, and the plot in Fig. 18.24 can be used to determine the maximum value permitted for $(W/L)_5/(W/L)_1$ while keeping $V_{\overline{Q}}$ below a desired value. Alternatively, we can derive a formula for this purpose. For instance, if $V_{\overline{Q}}$ is to be kept below V_{tn}, the ratio of $(W/L)_5$ to

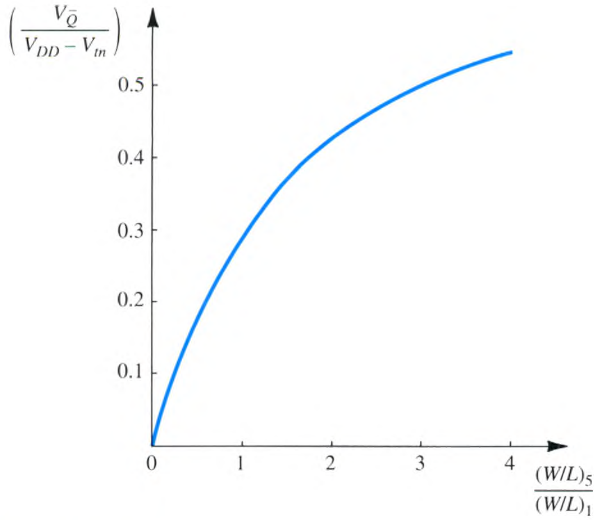


Figure 18.24 The normalized value of $V_{\bar{Q}}$ versus the ratio $(W/L)_5/(W/L)_1$ for the circuit in Fig. 18.23. This graph can be used to determine the maximum value permitted for $(W/L)_5/(W/L)_1$ so that $V_{\bar{Q}}$ is kept below a desired level.

$(W/L)_1$ must be kept below the value obtained from Eq. (18.21), that is,

$$\frac{(W/L)_5}{(W/L)_1} \leq \frac{1}{\left(1 - \frac{V_{tn}}{V_{DD} - V_{tn}}\right)^2} - 1 \qquad (18.22)$$

This is an important design constraint that can be expressed in a slightly more general form by replacing $(W/L)_5$ with $(W/L)_a$, where the subscript a denotes access transistors Q_5 and Q_6, and $(W/L)_1$ with $(W/L)_n$, which is the W/L ratio of Q_N in each of the two inverters; thus,

$$\frac{(W/L)_a}{(W/L)_n} \leq \frac{1}{\left(1 - \frac{V_{tn}}{V_{DD} - V_{tn}}\right)^2} - 1 \qquad (18.23)$$

EXERCISE

18.6 Find the maximum allowable W/L for the access transistors of the SRAM cell in Fig. 18.22 so that in a read operation, the voltages at Q and $\bar{Q}$ do not change by more than $|V_t|$. Assume that the SRAM is fabricated in a 65-nm technology for which $V_{DD} = 1.0$ V, $V_{tn} = |V_{tp}| = 0.35$ V and that $(W/L)_n = 2$.
Ans. $(W/L)_a \leq 7.4$

Having determined the constraint imposed by the read operation on the W/L ratios of the access transistors, we now return to the circuit in Fig. 18.23, and show in Fig. 18.25 the

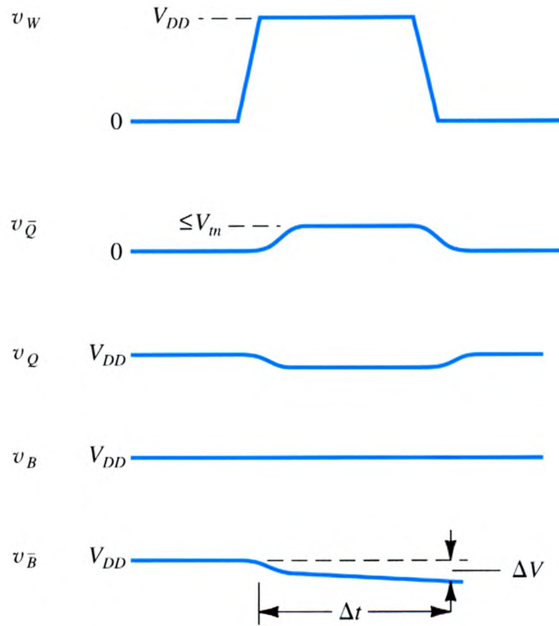


Figure 18.25 Voltage waveforms at various nodes in the SRAM cell during a read-1 operation.

voltage waveforms at various nodes during a read-1 operation. Observe that as we have already discussed, $v_{\overline{Q}}$ rises from zero to a voltage $V_{\overline{Q}} \leq V_{tn}$. Correspondingly, the change in v_Q will be very small, justifying the assumption implicit in the analysis above that v_Q remains constant at V_{DD}. Most important, note that the voltage of the $\overline{B}$ line, $v_{\overline{B}}$, decreases by a small amount ΔV. This is a result of the discharge of the capacitance of the $\overline{B}$ line, $C_{\overline{B}}$, by the current I_5. Assuming that I_5 reaches its equilibrium value in Eq. (18.19) relatively quickly, capacitor $C_{\overline{B}}$ is in effect discharged by a constant current I_5 and the change in its voltage, ΔV, obtained in a time interval Δt, can be found by writing a charge-balance equation,

$$I_5 \Delta t = C_{\overline{B}} \Delta V$$

Thus,

$$\Delta V = \frac{I_5 \Delta t}{C_{\overline{B}}} \tag{18.24}$$

Here we note that $C_{\overline{B}}$ is usually relatively large (possibly over 100 fF) because a large number of cells are connected to the $\overline{B}$ line. The incremental change ΔV is therefore rather small (0.1–0.2 V), necessitating the use of a sense amplifier. If the sense amplifier requires a minimum decrement ΔV in $v_{\overline{B}}$ to detect the presence of a "1," then the read delay time can be found from Eq. (18.24) as

$$\Delta t = \frac{C_{\overline{B}} \Delta V}{I_5} \tag{18.25}$$

This equation indicates the need for a relatively large I_5 to reduce the delay time Δt. A large I_5, however, implies selecting $(W/L)_a$ near the upper bound given by Eq. (18.23), which in turn means an increase in the silicon area occupied by the access transistors and hence the cell area, an interesting design trade-off.

EXERCISE

18.7 For the SRAM cell considered in Exercise 18.6, whose $(W/L)_n = 2$ and $(W/L)_a \leq 7.4$, use Eq. (18.25) to determine the read delay Δt in two cases: (a) $(W/L)_a = 7.4$ and (b) $(W/L)_a = 2$. Let $\mu_n C_{ox} = 450\ \mu A/V^2$. In both cases, assume that $C_{\overline{B}} = 2$ pF and that the sense amplifier requires a ΔV of minimum magnitude of 0.15 V. (*Hint*: Use Eq. 18.19 to determine I_5, and Eq. 18.21 to determine $V_{\overline{Q}}$.)

Ans. 2.0 ns; 3.2 ns

We conclude our discussion of the read operation with two remarks:

1. Although we considered only the read-1 operation, the read-0 operation is identical; it involves Q_3 and Q_6 with the analysis resulting in an upper bound on $(W/L)_6/(W/L)_3$ equal to what we have found for $(W/L)_5/(W/L)_1$. This, of course, is entirely expected, since the circuit is symmetrical. The read-0 operation results in a decrement ΔV in the voltage of the B line, which is interpreted by the sense amplifier as a stored 0.
2. The component Δt of the read delay is relatively large because C_B and $C_{\overline{B}}$ are relatively large (in the 100 fF range). Also, Δt is not the only component of the read delay; another significant component is due to the finite rise time of the voltage on the word line. Indeed, even the calculation of Δt is optimistic, since the word line will have only reached a voltage lower than V_{DD} when the process of discharging $C_{\overline{B}}$ takes place. As we will see shortly, the write operation is faster.

The Write Operation We next consider the write operation. Let the SRAM cell of Fig. 18.22 be storing a logic 1, thus $V_Q = V_{DD}$ and $V_{\overline{Q}} = 0$ V, and assume that we wish to write a 0; that is, we wish to have the flip-flop switch states. To write a zero, the B line is lowered to 0 V, and the $\overline{B}$ line is raised to V_{DD} and, of course, the cell is selected by raising the word line to V_{DD}. The objective now is to pull node Q down and node $\overline{Q}$ up and have the voltage of at least one of these two nodes pass by the inverter threshold voltage. Thus, if v_Q decreases below the threshold voltage of inverter Q_1–Q_2, the regenerative action of the latch will start and the flip-flop will switch to the stored-0 state. Alternatively, or in addition, if we manage to raise $v_{\overline{Q}}$ above the threshold voltage of the Q_3–Q_4 inverter, the regenerative action will be engaged and the latch will eventually switch state. Either one of the two actions is sufficient to engage the regenerative mechanism of the latch.

Figure 18.26 shows the relevant parts of the SRAM circuit during the interval when $v_{\overline{Q}}$ is being pulled up [Fig. 18.26(a)] and v_Q is being pulled down [Fig. 18.26(b)]. Since **toggling** (i.e., state change) has not yet taken place, we assume that the voltage feeding the gate of Q_1 is still equal to V_{DD} and the voltage at the gate of Q_4 is still equal to 0 V. These voltages will of course be changing as $v_{\overline{Q}}$ goes up and v_Q goes down, but this assumption is nevertheless reasonable for approximate hand analysis.

Consider first the circuit in Fig. 18.26(a). This is the same circuit we analyzed in detail in the study of the read operation above. Recall that to make the read process nondestructive, we imposed an upper bound on $(W/L)_5$. That upper bound ensured that $v_{\overline{Q}}$ will not rise above V_{tn}. Thus, this circuit is not capable of raising $v_{\overline{Q}}$ to the point that it can start the regenerative action. We must therefore rely solely on the circuit of Fig. 18.26(b). That is, our write-0

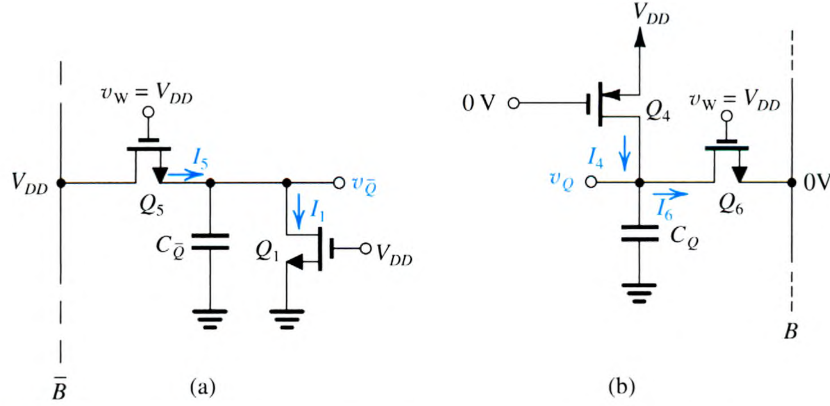


Figure 18.26 Relevant parts of the 6T SRAM circuit of Fig. 18.22 during the process of writing a 0. It is assumed that the cell is originally storing a 1 and thus initially $v_Q = V_{DD}$ and $v_{\bar{Q}} = 0$ V.

operation will be accomplished by pulling node Q down in order to initiate the regenerative action of the latch. To ensure that the latch will in fact switch state, SRAM designers impose a more stringent requirement on the voltage v_Q, namely, that it must fall below not just V_M of the Q_1–Q_2 inverter but even further such that it is lower than V_{tn} of Q_1.

Let's now look more closely at the circuit of Fig. 18.26(b). Initially, v_Q is at V_{DD}. However, as Q_6 turns on, I_6 quickly discharges the small capacitance C_Q, and v_Q begins to fall. This will enable Q_4 to conduct, and equilibrium is reached when $I_4 = I_6$. To ensure toggling, we design the circuit so that this equilibrium occurs at a value of v_Q less than V_{tn}. At such a value V_Q, Q_4 will be operating in saturation (or at least at the edge of saturation) and Q_6 will be operating in the triode region, thus

$$I_4 = \frac{1}{2}(\mu_p C_{ox})\left(\frac{W}{L}\right)_4 (V_{DD} - |V_{tp}|)^2 \tag{18.26}$$

and

$$I_6 = (\mu_n C_{ox})\left(\frac{W}{L}\right)_6 \left[(V_{DD} - V_{tn})V_Q - \frac{1}{2}V_Q^2\right] \tag{18.27}$$

Substituting $|V_{tp}| = V_{tn}$, which is usually the case, and equating I_4 and I_6 results in a quadratic equation in V_Q whose solution is

$$\frac{V_Q}{V_{DD} - V_{tn}} = 1 - \sqrt{1 - \left(\frac{\mu_p}{\mu_n}\right)\frac{(W/L)_4}{(W/L)_6}} \tag{18.28}$$

This relationship is not as convenient as the one in Eq. (18.21) because the right-hand side includes a process-dependent quantity, namely, μ_p/μ_n. Thus we do not have a universally applicable relationship. Nevertheless, for a number of CMOS process technologies, $\mu_n/\mu_p \simeq 4$.[2] Thus, upon substituting $\mu_p/\mu_n = 0.25$ in Eq. (18.28), we obtain the semiuniversal graph shown in Fig. 18.27. We can use this graph to determine the maximum allowable value of the ratio $(W/L)_4/(W/L)_6$ that will ensure a value of $V_Q \leq V_{tn}$ for given process parameters

[2] In recent nanoscale technologies (e.g., 28-nm CMOS), this ratio is closer to 1.

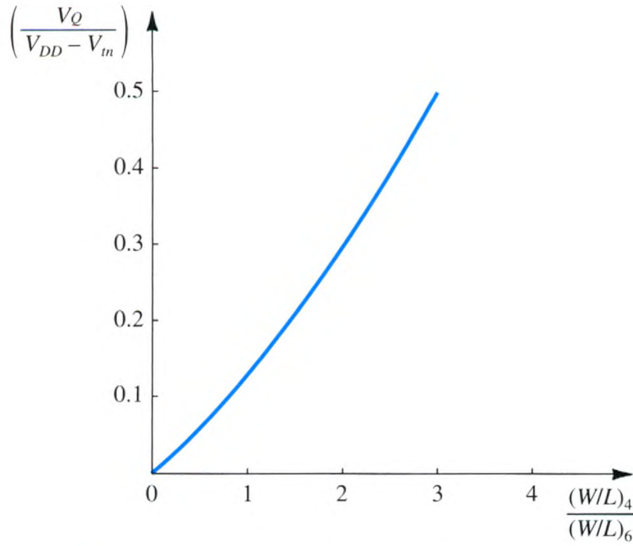


Figure 18.27 The normalized value of V_Q versus the ratio $(W/L)_4/(W/L)_6$ for the circuit in Fig. 18.26(b). The graph applies for process technologies for which $\mu_n \simeq 4\mu_p$. It can be used to determine the maximum $(W/L)_4/(W/L)_6$ for which V_Q is guaranteed to fall below a desired value.

V_{DD} and V_{tn}. Alternatively, substituting $V_Q = V_{tn}$, $(W/L)_4 = (W/L)_p$, and $(W/L)_6 = (W/L)_a$, we can obtain the upper bound analytically as

$$\frac{(W/L)_p}{(W/L)_a} \leq \left(\frac{\mu_n}{\mu_p}\right)\left[1 - \left(1 - \frac{V_{tn}}{V_{DD} - V_{tn}}\right)^2\right] \tag{18.29}$$

Observe that this relationship provides an upper bound on $(W/L)_p$ in terms of $(W/L)_a$ and that the relationship in Eq. (18.23) provides an upper bound on $(W/L)_a$ in terms of $(W/L)_n$. Thus, the two relationships can be used together to design the SRAM cell.

EXERCISE

D18.8 For the SRAM cell considered in Exercise 18.6, where $(W/L)_n = 2$ and $(W/L)_a = 7.4$, use Eq. (18.29) to find the maximum allowable value of $(W/L)_p$. For this 65-nm process, you may consider that $\mu_n \simeq 3\mu_p$. For all transistors having $L = 65$ nm, find W_n, W_p, and W_a that result in a minimum-area cell. Assume that the minimum allowable width is 130 nm.

Ans. $(W/L)_p \leq 2.36 \times (W/L)_a$, thus $(W/L)_p \leq 17.5$; for minimum area, select $W_n = W_p = W_a = 130$ nm.

We conclude our study of the write process by noting that it is fast because it does not require discharging the large capacitance of the bit lines. The voltages of the B and $\overline{B}$ lines are driven to their required values of 0 or V_{DD} by powerful driver circuits and thus achieve

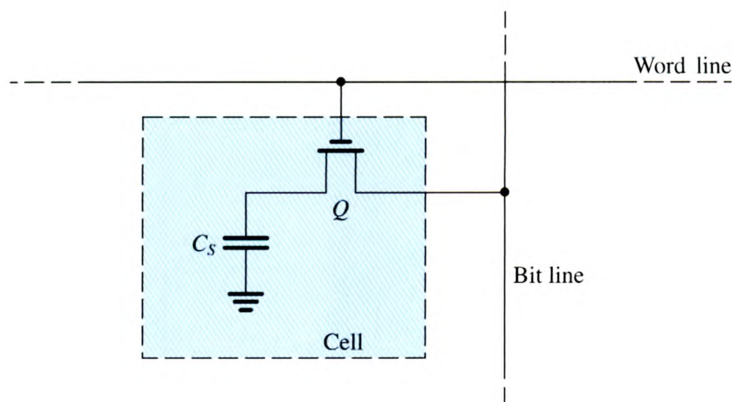


Figure 18.28 The one-transistor dynamic RAM (DRAM) cell.

their desired voltages very quickly. The write delay is determined roughly by the time for the regenerating signal to propagate around the feedback loop of the latch; thus it is about twice the propagation delay of the inverter. Of course, the write cycle time is still lengthened by the word-line delay.

18.3.2 Dynamic Memory (DRAM) Cell

Although a variety of DRAM storage cells have been proposed over the years, a particular cell, shown in Fig. 18.28, has become the industry standard. The cell consists of a single *n*-channel MOSFET, known as the **access transistor**, and a **storage capacitor** C_S. The cell is appropriately known as the **one-transistor cell**.[3] The gate of the transistor is connected to the word line, and its source (drain) is connected to the bit line. Observe that only one bit line is used in DRAMs, whereas in SRAMs both the bit and $\overline{\text{bit}}$ lines are utilized. The DRAM cell stores its bit of information as charge on the cell capacitor C_S. When the cell is storing a 1, the capacitor is charged to V_{DD}; when a 0 is stored, the capacitor is discharged to zero volts.

Some explanation is needed to appreciate how the capacitor can be charged to the full supply voltage V_{DD}. Consider a write-1 operation. The word line is at V_{DD} and the bit line is at V_{DD} and the transistor is conducting, charging C_S. The transistor will cease conduction when the voltage on C_S reaches $(V_{DD} - V_t)$. This is the same problem we encountered with the NMOS pass transistor in Section 18.1. The problem is overcome in DRAM design by boosting the word line to a voltage equal to $V_{DD} + V_t$. In this case the capacitor voltage for a stored 1 will be equal to the full V_{DD}. However, because of leakage effects, the capacitor charge will leak off, and hence the cell must be refreshed periodically. During **refresh**, the cell content is read and the data bit is rewritten, thus *restoring* the capacitor voltage to its proper value. Typically, the refresh operation must be performed every 10 ms to 100 ms.

Let us now consider the DRAM operation in more detail. As in the static RAM, the row decoder selects a particular row by raising the voltage of its word line. This causes all the access transistors in the selected row to become conductive, thereby connecting the storage capacitors of all the cells in the selected row to their respective bit lines. Thus the cell capacitor C_S is connected in parallel with the bit-line capacitance C_B, as indicated in

[3]The name was originally used to distinguish this cell from earlier ones utilizing three transistors.



Figure 18.29 When the voltage of the selected word line is raised, the transistor conducts, thus connecting the storage capacitor C_S to the bit-line capacitor C_B.

Fig. 18.29. Here, it should be noted that C_S is typically 20 fF to 30 fF, whereas C_B is 10 times larger. Now, if the operation is a read, the bit line is precharged to $V_{DD}/2$. To find the change in the voltage on the bit line resulting from connecting a cell capacitor C_S to it, let the initial voltage on the cell capacitor be V_{CS} ($V_{CS} = V_{DD}$ when a 1 is stored, and $V_{CS} = 0$ V when a 0 is stored). Using charge conservation, we can write

$$C_S V_{CS} + C_B \frac{V_{DD}}{2} = (C_B + C_S)\left(\frac{V_{DD}}{2} + \Delta V\right)$$

from which we can obtain for ΔV

$$\Delta V = \frac{C_S}{C_B + C_S}\left(V_{CS} - \frac{V_{DD}}{2}\right) \tag{18.30}$$

and since $C_B \gg C_S$,

$$\Delta V \simeq \frac{C_S}{C_B}\left(V_{CS} - \frac{V_{DD}}{2}\right) \tag{18.31}$$

Now, if the cell is storing a 1, $V_{CS} = V_{DD}$, and

$$\Delta V(1) \simeq \frac{C_S}{C_B}\left(\frac{V_{DD}}{2}\right) \tag{18.32}$$

whereas if the cell is storing a 0, $V_{CS} = 0$, and

$$\Delta V(0) \simeq -\frac{C_S}{C_B}\left(\frac{V_{DD}}{2}\right) \tag{18.33}$$

Since usually C_B is much greater than C_S, these readout voltages are very small. For example, for $C_B = 10C_S$, $V_{DD} = 1.8$ V, $\Delta V(0)$ will be about –90 mV, and $\Delta V(1)$ will be +90 mV. This is a best-case scenario, for the 1 level in the cell might very well be below V_{DD}. Furthermore, in modern memory chips, V_{DD} is 1.2 V or even lower. In any case, we see that a stored 1 in the cell results in a small positive increment in the bit-line voltage, whereas a stored zero results in a small negative increment. Observe also that the readout process is *destructive,* since the resulting voltage across C_S will no longer be V_{DD} or 0.

The change of voltage on the bit line is detected and amplified by the column sense amplifier, causing the bit line to be driven to the full-scale value (0 or V_{DD}) of the detected signal. This amplified signal is then impressed on the storage capacitor, thus restoring its signal to the proper level (V_{DD} or 0). In this way, all the cells in the selected row are refreshed. Simultaneously, the signal at the output of the sense amplifier of the selected column is fed to the data-output line of the chip through the action of the column decoder.

The write operation proceeds similarly to the read operation, except that the data bit to be written, which is impressed on the data-input line, is applied by the column decoder to the selected bit line. Thus, if the data bit to be written is a 1, the bit-line voltage is raised to V_{DD}

(i.e., C_B is charged to V_{DD}). When the access transistor of the particular cell is turned on, its capacitor C_S will be charged to V_{DD}; thus a 1 is written in the cell. Simultaneously, all the other cells in the selected row are simply refreshed.

Although the read and write operations result in automatic refreshing of all the cells in the selected row, provision must be made for the periodic refreshing of the entire memory, typically every 10 ms to 100 ms, as specified for the particular chip. The refresh operation is carried out in a *burst mode,* one row at a time. During refresh, the chip will not be available for read or write operations. This is not a serious matter, however, since the interval required to refresh the entire chip is typically less than 2% of the time between refresh cycles. In other words, the memory chip is available for normal operation more than 98% of the time.

EXERCISE

18.9 In a particular dynamic memory chip, $C_S = 30$ fF, $C_B = 0.3$ pF, and $V_{DD} = 1.2$ V. Find the output readout voltage for a stored 1 and a stored 0. Recall that in a read operation, the bit lines are precharged to $V_{DD}/2$.
Ans. 60 mV; –60 mV

18.3.3 Flash Memory

Flash memory is an important type of nonvolatile memory used in solid-state storage devices as well as in USB keys and memory cards.

Flash memories use variants of the memory cell whose cross section is shown in Fig. 18.30(a). The cell is basically an n-channel MOSFET with two gates instead of one, stacked on top of each other. One of the gates is not electrically connected to any other part of the circuit; rather, it is left floating and is appropriately called a **floating gate**. The other gate, called a **select gate**, functions in the same manner as the gate of a regular MOSFET. The MOS transistor of Fig. 18.30(a) is known as a **floating-gate transistor** and is given the circuit symbol shown in Fig. 18.30(b). In this symbol, the broken line denotes the floating gate.

A floating gate transistor can be configured into one of many states by trapping electrons in the floating gate, which is insulated. The amount of charge trapped in the floating gate impacts the threshold voltage of the transistor, and this can be detected during a read operation. Electrons are trapped by applying a large voltage between the drain and source, and a large voltage at the select gate. The drain-to-source voltage accelerates electrons through the channel. As these electrons reach the drain end of the channel, they acquire high kinetic energy and are referred to as *hot electrons*. The large positive voltage on the select gate (greater than the drain voltage) establishes an electric field in the insulating oxide. This electric field attracts the hot electrons and accelerates them (through the oxide) toward the floating gate. In this way the floating gate is charged, and the charge that accumulates on it becomes trapped. Fortunately, the process of charging the floating gate is self-limiting. The negative charge that accumulates on the floating gate reduces the strength of the electric

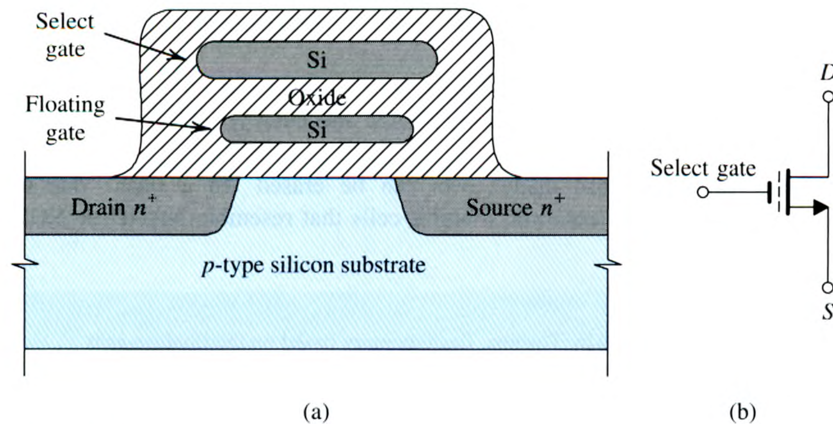


Figure 18.30 **(a)** Cross section and **(b)** circuit symbol of the floating-gate transistor used as a Flash cell.

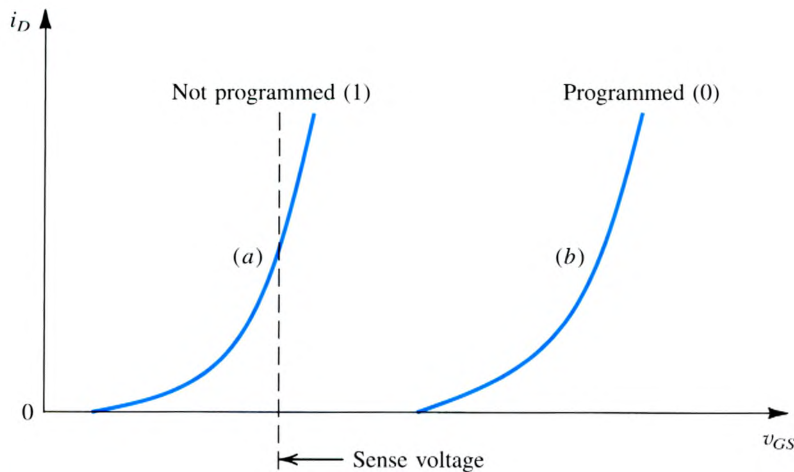


Figure 18.31 Illustrating the shift in the $i_D - v_{GS}$ characteristic of a floating-gate transistor as a result of programming.

field in the oxide to the point that it eventually becomes incapable of accelerating any more of the hot electrons.

The negative charge trapped on the floating gate will cause electrons to be repelled from the surface of the substrate. This implies that to form a channel, the positive voltage that has to be applied to the select gate will have to be greater than that required when the floating gate is not charged. In other words, the threshold voltage V_t of the programmed transistor (say, storing a logic 0) will be higher than that of the not-programmed (logic 1) device. This is illustrated in Fig. 18.31, which shows the $i_D - v_{GS}$ characteristics of (a) a not-programmed device and (b) a programmed device.

Reading the content of the floating-gate cell is easy: A voltage V_{GS} somewhere between the threshold values for logic 0 and logic 1 is applied to the selected gate. While a

programmed device (one that is storing a 0) will not conduct, a not-programmed device (one that is storing a 1) will conduct heavily.

The charge trapped on the floating gate, and consequently the stored state, persists even when the memory is powered off. Thus, unlike the SRAM and DRAM cells studied earlier in this section, a flash memory is nonvolatile.

The name "flash" arises because many rows can be erased "in a flash" due to organizations of floating-gate transistors into multibit cells that resemble NAND or NOR structures.

BLINDING FLASH

Since its invention in 1980 by Toshiba, flash memory based on the floating-gate MOS transistor has expanded into every possible field of computing. Because of its nonvolatility, flash memory has become largely responsible for the dominance of mobile digital devices. Increasingly, flash-based solid-state drives (SSDs) are overtaking hard disk drives in enterprise memory systems. SSDs of more than 1 terabyte with no moving parts are becoming available for a few hundred dollars in technologies as small as 20 nm. In many applications, high data rates of up to 12 Gb/s allow total replacement of volatile DRAMs in handheld devices. At the other end of the scale, flash-filled USB drives with gigabyte capacities have effectively replaced the need for DVDs in today's laptops.

18.4 Ring Oscillators and Special-Purpose Circuits

Having studied the circuits commonly used to implement the storage cells in SRAMs and DRAMs, we now consider some of the other important circuit blocks in a memory chip. The design of these circuits and other sequential logic circuits presents exciting challenges and opportunities to integrated-circuit designers: Improving the performance of peripheral circuits can result in denser and faster memory chips that dissipate less power, and can provide the necessary functions for high-speed operation in digital systems.

18.4.1 Ring Oscillators and Other Pulse-Generation Circuits

Memory chips and other sequential circuits require a large number of pulse signals, sometimes with intricate timing relationships among them. It is not our purpose here to study this important subject; rather, we present two simple circuits that find widespread application in memory-chip timing as well as in other digital system components, such as microprocessors.

The Ring Oscillator A ring oscillator is formed by connecting an odd number of inverters in a loop. Although usually at least five inverters are used, we illustrate the principle of operation using a ring of three inverters, as shown in Fig. 18.32(a). Figure 18.32(b) shows the waveforms obtained at the outputs of the three inverters. These waveforms are idealized in the sense that their edges have zero rise and fall times. Nevertheless, they will serve to explain the circuit operation.

Observe that a rising edge at node 1 propagates through gates 1, 2, and 3 to return inverted after a delay of $3t_P$; note that this represents a form of *negative* feedback. This falling edge then propagates, and returns with the original (rising) polarity after another $3t_P$ interval. It follows that the circuit oscillates with a period of $6t_P$ or, correspondingly, with frequency

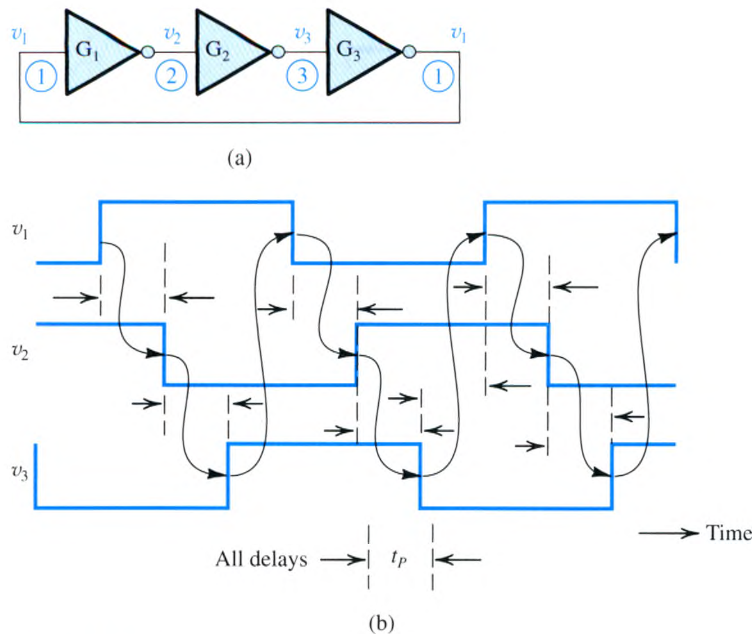


Figure 18.32 **(a)** A ring oscillator formed by connecting three inverters in cascade. (Normally at least five inverters are used.) **(b)** The resulting waveform. Observe that the circuit oscillates with frequency $1/6t_P$.

$1/6t_P$. In general, a ring with N inverters will oscillate with a period of $2Nt_P$ and frequency

$$f = 1/(2Nt_P) \tag{18.34}$$

An odd number of inverters, with $N \geq 3$, must be used for the circuit to oscillate. Using an even number of inverters results in positive feedback, where logic values are reinforced (notice that an SRAM cell uses $N = 2$). Furthermore, a ring with only one inverter does not oscillate since its feedback loop has only one node and thus only one pole. It will settle to the inverter threshold where its input and output voltages are equal.

Recall from Chapter 17 [Eqs. (17.8) and (17.10)] that t_P is inversely proportional to V_{DD}. Hence, we can control the frequency of a ring oscillator by adjusting the supply voltage. The resulting **voltage-controlled oscillator (VCO)** is widely used to generate clocking signals for integrated circuits. Finally, a ring oscillator provides a means for measuring the inverter propagation delay: simply measure the oscillation frequency, and then divide the period by $2N$.

EXERCISE

18.10 Find the frequency of oscillation of a ring of five inverters if the inverter propagation delay is 100 ps. If V_{DD} is reduced by 10%, what do you expect the oscillation frequency to become?
Ans. 1 GHz; 900 MHz

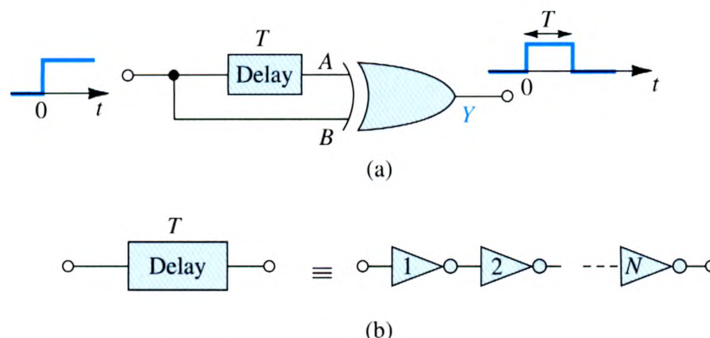


Figure 18.33 **(a)** A one-shot or monostable circuit. Utilizing a delay circuit with a delay T and an XOR gate, this circuit provides an output pulse of width T. **(b)** The delay circuit can be implemented as the cascade of N inverters where N is even, in which case $T = Nt_p$.

A One-Shot or Monostable Multivibrator Circuit A one-shot or monostable multivibrator circuit provides, when triggered, a single output pulse with a predetermined width.[4] In Fig. 18.33(a), we show a commonly used circuit that utilizes an exclusive-OR (XOR) gate together with a delay element. Recalling that the XOR gate provides a high output only when its two inputs are dissimilar, we see that prior to the arrival of the input positive step, the output will be low.

When the input goes high, only the B input of the XOR will be high and thus its output will go high. The high input will reach input A of the XOR T seconds later, at which time both inputs of the XOR will be high and thus its output will go low. We thus see that the circuit produces an output pulse with a duration T equal to the delay of the delay block for each transition of the input signal. The delay block can be implemented by connecting an even number of inverters in cascade as shown in Fig. 18.33(b).

Monostable circuits are used in asynchronous systems where short pulses signal to other circuits that a value has been updated. Also, the output of a monostable circuit can be used to get a latch to act like a flip-flop; here, the short pulse is used as a clock input, giving the latch an edge-triggered behavior.

18.4.2 The Sense Amplifier

Another important peripheral circuit that is used in memory chips is the sense amplifier. Next to the storage cells, the sense amplifier is the most critical component in a memory chip. Sense amplifiers are essential to the proper operation of DRAMs, and their use in SRAMs results in speed and area improvements.

A variety of sense-amplifier designs are in use, some of which closely resemble the active-load MOS differential amplifier studied in Chapter 9. Here, we describe a differential sense amplifier that employs positive feedback. Because the circuit is differential, it can be employed directly in SRAMs, where the SRAM cell utilizes both the B and $\overline{B}$ lines. On the

[4]The name "monostable" arises because this class of circuits has one stable state, which is the quiescent state. When a trigger is applied, the circuit moves to its quasi-stable state and stays in it for a predetermined length of time (the width of the output pulse). It then switches back automatically to the stable state.

other hand, the one-transistor DRAM circuit we studied in Section 18.3.2 is a single-ended circuit, utilizing one bit line only. The DRAM circuit, however, is made to resemble a differential signal source by comparing its output voltage to that generated by a "dummy cell" where a voltage halfway between a logic '1' and '0' is stored. Therefore, we shall assume that the memory cell whose output is to be amplified develops a difference output voltage between the B and $\overline{B}$ lines. This signal, which can range from 20 mV to 500 mV depending on the memory type and cell design, will be applied to the input terminals of the sense amplifier. The sense amplifier in turn responds by providing a full-swing (0 to V_{DD}) signal at its output terminals. The particular amplifier circuit we shall discuss here has a rather unusual property: *Its output and input terminals are the same!*

A Sense Amplifier with Positive Feedback Figure 18.34 shows the sense amplifier together with some of the other column circuitry of a RAM chip. Note that the sense amplifier is nothing but the familiar latch formed by cross-coupling two CMOS inverters: One inverter is implemented by transistors Q_1 and Q_2, and the other by transistors Q_3 and Q_4. Transistors Q_5 and Q_6 act as switches that connect the sense amplifier to ground and V_{DD} only when data-sensing action is required. Otherwise, ϕ_s is low and the sense amplifier is turned off. This conserves power, an important consideration because usually there is one sense amplifier per column, resulting in *thousands of sense amplifiers per chip*. Note, again, that terminals x and y are both the input and the output terminals of the amplifier. As indicated, these I/O terminals are connected to the B and $\overline{B}$ lines. The amplifier is required to detect a small signal appearing between B and $\overline{B}$, and to amplify it to provide a full-swing signal at B and $\overline{B}$. For instance, if during a read operation, the cell has a stored 1, then a small positive voltage will develop between B and $\overline{B}$, with v_B higher than $v_{\overline{B}}$. The amplifier will then cause v_B to rise to V_{DD} and $v_{\overline{B}}$ to fall to 0 V. This 1 output is then directed to the chip I/O pin by the column decoder (not shown) and at the same time is used to rewrite a 1 in the DRAM cell, thus performing the restore operation that is required because the DRAM readout process is destructive.

Figure 18.34 also shows the precharge and equalization circuit. Operation of this circuit is straightforward: When ϕ_p goes high (to V_{DD}) prior to a read operation, all three transistors conduct. While Q_8 and Q_9 precharge the $\overline{B}$ and B lines to $V_{DD}/2$, transistor Q_7 helps speed up this process by equalizing the initial voltages on the two lines. This equalization is critical to the proper operation of the sense amplifier. Any voltage difference present between B and $\overline{B}$ prior to commencement of the read operation can result in erroneous interpretation by the sense amplifier of its input signal. In Fig. 18.34, we show only one of the cells in this particular column, namely, the cell whose word line is activated. The cell can be either an SRAM or a DRAM cell. All other cells in this column will not be connected to the B and $\overline{B}$ lines (because their word lines will remain low).

Let us now consider the sequence of events during a read operation:

1. The precharge and equalization circuit is activated by raising the control signal ϕ_p. This will cause the B and $\overline{B}$ lines to be at equal voltages, equal to $V_{DD}/2$. The clock ϕ_p then goes low, and the B and $\overline{B}$ lines are left to float for a brief interval.
2. The word line goes up, connecting the cell to the B and $\overline{B}$ lines. A voltage then develops between B and $\overline{B}$, with v_B higher than $v_{\overline{B}}$ if the accessed cell is storing a 1, or v_B lower than $v_{\overline{B}}$ if the cell is storing a 0. To keep the cell area small, and to facilitate operation at higher speeds, the readout signal, which the cell is required to provide between B and $\overline{B}$, is kept small (typically, 20–500 mV).

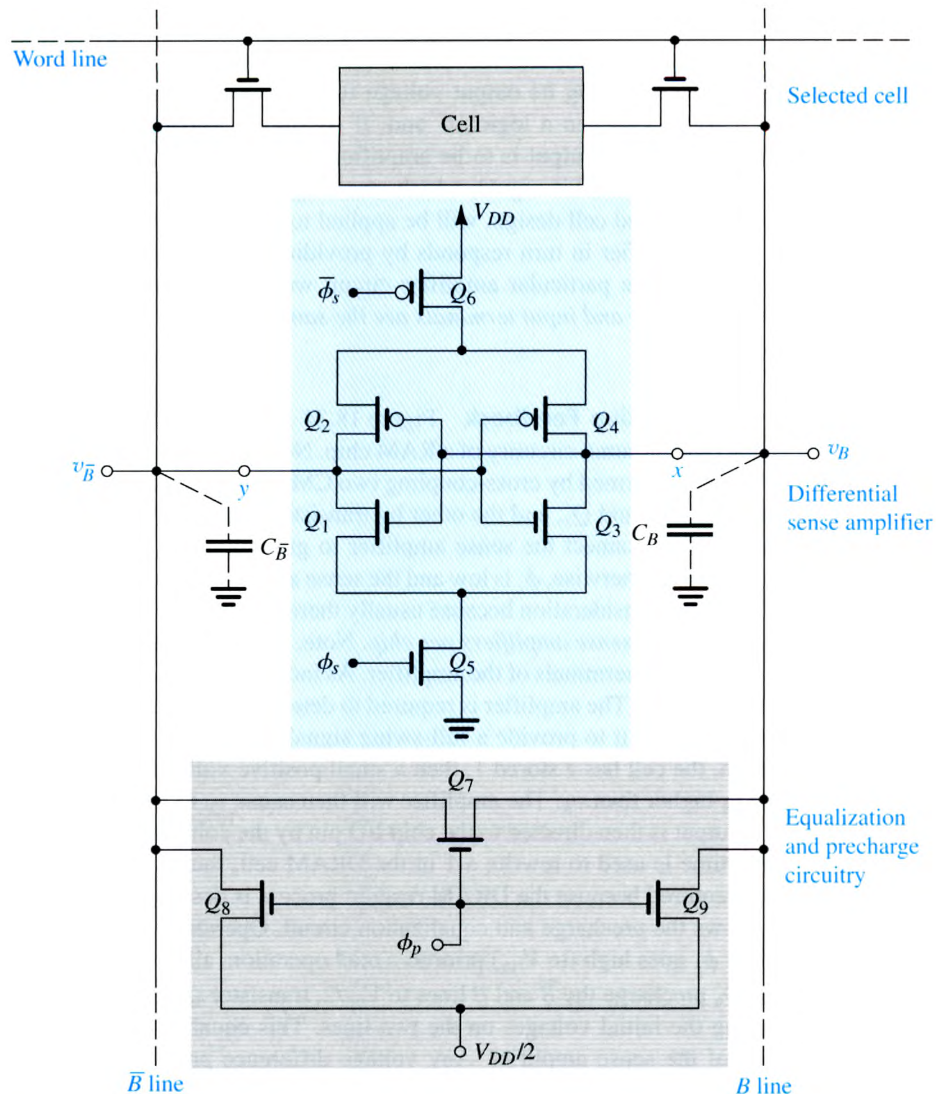


Figure 18.34 A differential sense amplifier connected to the bit lines of a particular column. This arrangement can be used directly for SRAMs (which utilize both the B and $\overline{B}$ lines). DRAMs can be turned into differential circuits by using a "dummy-cell" arrangement.

3. Once an adequate difference voltage signal between B and $\overline{B}$ has been developed by the storage cell, the sense amplifier is turned on by connecting it to ground and V_{DD} through Q_5 and Q_6, activated by raising the sense-control signal ϕ_s. Because initially the input terminals of the inverters are at $V_{DD}/2$, the inverters will be operating in their transition region, where the gain is high (Section 16.3). It follows that initially the latch will be operating at its unstable equilibrium point. Thus, depending on the signal between the input terminals, the latch will quickly move to one of its two stable equilibrium points (refer to the description of the latch operation in Section 18.2). This is achieved by the regenerative action, inherent in positive

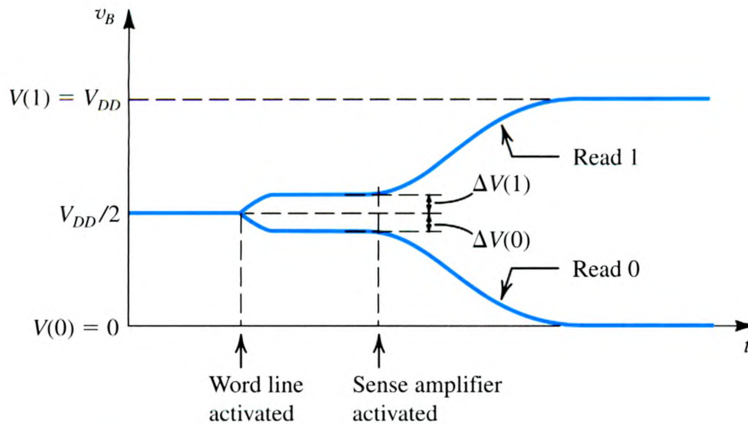


Figure 18.35 Waveforms of v_B before and after the activation of the sense amplifier. In a read-1 operation, the sense amplifier causes the initial small increment $\Delta V(1)$ to grow exponentially to V_{DD}. In a read-0 operation, the negative $\Delta V(0)$ grows to 0. Complementary signal waveforms develop on the $\overline{B}$ line.

feedback. Figure 18.35 clearly illustrates this point by showing the waveforms of the signal on the bit line for both a read-1 and a read-0 operation. Observe that once activated, the sense amplifier causes the small initial difference, $\Delta V(1)$ or $\Delta V(0)$, provided by the cell, to grow exponentially to either V_{DD} (for a read-1 operation) or 0 (for read-0 operation). The waveforms of the signal on the $\overline{B}$ line will be complementary to those shown in Fig. 18.35 for the B line. In the following, we quantify the process of exponential growth of v_B and $v_{\overline{B}}$.

A Closer Look at the Operation of the Sense Amplifier Developing a precise expression for the output signal of the sense amplifier shown in Fig. 18.34 is a rather complex task requiring the use of large-signal (and thus nonlinear) models of the inverter voltage-transfer characteristic, as well as taking the positive feedback into account. We will not do this here; rather, we shall consider the operation in a semiquantitative way.

Recall that at the time the sense amplifier is activated, each of its two inverters is operating in the transition region near $V_{DD}/2$. Thus, for small-signal operation, each inverter can be modeled using g_{mn} and g_{mp}, the transconductances of Q_N and Q_P, respectively, evaluated at an input bias of $V_{DD}/2$. Specifically, a small-signal v_i superimposed on $V_{DD}/2$ at the input of one of the inverters gives rise to an inverter output current signal of $(g_{mn} + g_{mp})v_i \equiv G_m v_i$. This output current is delivered to one of the capacitors, C_B or $C_{\overline{B}}$. The voltage thus developed across the capacitor is then fed back to the other inverter and is multiplied by its G_m, which gives rise to an output current feeding the other capacitor, and so on, in a regenerative process. The positive feedback in this loop will mean that the signal around the loop, and thus v_B and $v_{\overline{B}}$, will *rise or decay exponentially* (see Fig. 18.35) with a time constant of (C_B/G_m) [or $(C_{\overline{B}}/G_m)$, since we have been assuming $C_B = C_{\overline{B}}$]. Thus, for example, in a read-1 operation we obtain

$$v_B = \frac{V_{DD}}{2} + \Delta V(1)e^{(G_m/C_B)t}, \qquad v_B \leq V_{DD} \tag{18.35}$$

whereas in a read-0 operation,

$$v_B = \frac{V_{DD}}{2} - \Delta V(0)e^{(G_m/C_B)t}, \qquad v_B \geq 0 \tag{18.36}$$

Because these expressions have been derived assuming small-signal operation, they describe the exponential growth (decay) of v_B reasonably accurately only for values close to $V_{DD}/2$. Nevertheless, they can be used to obtain a reasonable estimate of the time required to develop a particular signal level on the bit line.

Example 18.3

Consider the sense-amplifier circuit of Fig. 18.34 during the reading of a 1. Assume that the storage cell provides a voltage increment on the B line of $\Delta V(1) = 0.1$ V. If the NMOS devices in the amplifiers have $(W/L)_n = 0.54\ \mu\text{m}/0.18\ \mu\text{m}$ and the PMOS devices have $(W/L)_p = 2.16\ \mu\text{m}/0.18\ \mu\text{m}$, and assuming that $V_{DD} = 1.8$ V, $V_{tn} = |V_{tp}| = 0.5$ V, and $\mu_n C_{ox} = 4\ \mu_p C_{ox} = 300\ \mu\text{A/V}^2$, find the time required for v_B to reach $0.9\ V_{DD}$. Assume $C_B = 1$ pF.

Solution

First, we determine the transconductances g_{mn} and g_{mp}

$$g_{mn} = \mu_n C_{ox}\left(\frac{W}{L}\right)_n (V_{GS} - V_t)$$

$$= 300 \times \frac{0.54}{0.18}(0.9 - 0.5)$$

$$= 0.36 \text{ mA/V}$$

$$g_{mp} = \mu_p C_{ox}\left(\frac{W}{L}\right)_p (V_{SG} - |V_t|)$$

$$= 75 \times \frac{2.16}{0.18}(0.9 - 0.5) = 0.36 \text{ mA/V}$$

Thus, the inverter G_m is

$$G_m = g_{mn} + g_{mp} = 0.72 \text{ mA/V}$$

and the time constant τ for the exponential growth of v_B will be

$$\tau \equiv \frac{C_B}{G_m} = \frac{1 \times 10^{-12}}{0.72 \times 10^{-3}} = 1.4 \text{ ns}$$

Now, the time, Δt, for v_B to reach 0.9 V_{DD} can be determined from

$$0.9 \times 1.8 = 0.9 + 0.1e^{\Delta t/1.4}$$

resulting in

$$\Delta t = 2.8 \text{ ns}$$

EXERCISE

18.11 Repeat Example 18.3 for a sense-amplifier circuit designed in a 65-nm process where V_{DD} = 1.0 V, $V_{tn} = |V_{tp}| = 0.35$ V, $\mu_n C_{ox} = 3\mu_p C_{ox} = 450$ μA/V^2, and where the bit line capacitance is reduced to C_B = 0.25 pF. Use the following transistor dimensions: $(W/L)_n$ = 130 nm/65 nm and $(W/L)_p$ = 390 nm/65 nm.
Ans. 1.3 ns

18.4.3 The Row-Address Decoder

As described in Section 18.3, the row-address decoder is required to select one of the 2^M word lines in response to an M-bit address input. As an example, consider the case $M = 3$ and denote the three address bits A_0, A_1, and A_2, and the eight word lines $W_0, W_1, \ldots, W_7$. Conventionally, word line W_0 will be high when $A_0 = 0$, $A_1 = 0$, and $A_2 = 0$; thus we can express W_0 as a Boolean function of A_0, A_1, and A_2,

$$W_0 = \overline{A}_0\overline{A}_1\overline{A}_2 = \overline{A_0 + A_1 + A_2}$$

Thus the selection of W_0 can be accomplished by a three-input NOR gate whose three inputs are connected to A_0, A_1, and A_2 and whose output is connected to word line 0. Word line W_3 will be high when $A_0 = 1$, $A_1 = 1$, and $A_2 = 0$; thus,

$$W_3 = A_0 A_1 \overline{A}_2 = \overline{\overline{A}_0 + \overline{A}_1 + A_2}$$

Thus the selection of W_3 can be realized by a three-input NOR gate whose three inputs are connected to $\overline{A}_0$, $\overline{A}_1$, and A_2, and whose output is connected to word line 3. We can thus see that this address decoder can be realized by eight three-input NOR gates. Each NOR gate is fed with the appropriate combination of address bits and their complements, corresponding to the word line to which its output is connected.

A simple approach to realizing these NOR functions is provided by the matrix structure shown in Fig. 18.36. Attached to each row line is a p-channel device that is activated, prior to the decoding process, using the precharge control signal ϕ_P. During precharge

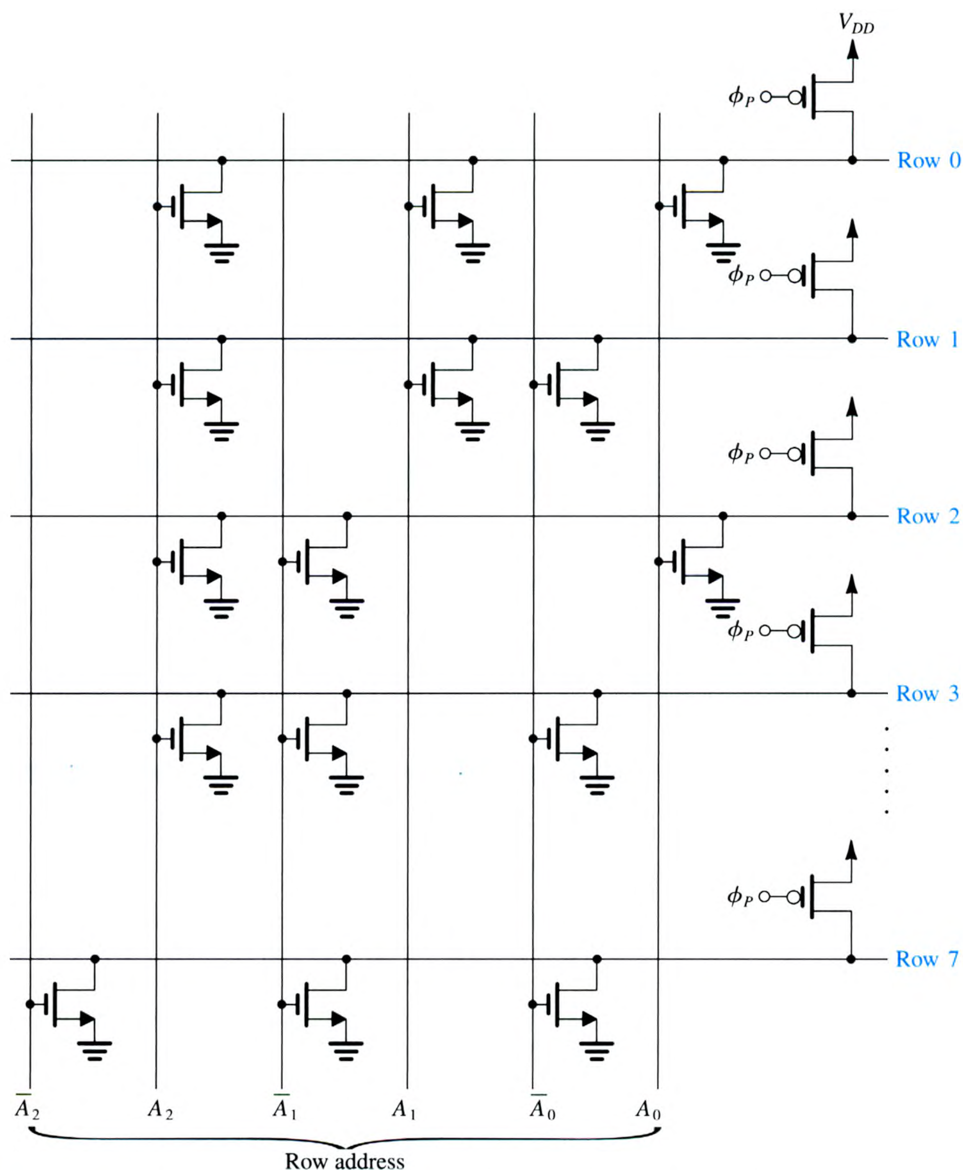


Figure 18.36 A NOR address decoder in array form. One out of eight lines (row lines) is selected using a 3-bit address.

(ϕ_P low), all the word lines are pulled high to V_{DD}. It is assumed that at this time the address input bits have not yet been applied and all the inputs are low. Then, the decoding operation begins when the address bits and their complements are applied. Observe that the NMOS transistors are placed so that the word lines not selected will be discharged. For any input combination, only one word line will not be discharged, and thus its voltage remains high at V_{DD}. For instance, row 0 will be high only when $A_0 = 0$, $A_1 = 0$, and $A_2 = 0$; this is the only combination that will result in all three transistors connected to row 0 being cut off. Similarly, row 3 has transistors connected to $\overline{A}_0, \overline{A}_1,$

and A_2, and thus it will be high when $A_0 = 1$, $A_1 = 1$, $A_2 = 0$, and so on. After the decoder outputs have stabilized, the output lines are connected to the word lines of the array, usually via clock-controlled transmission gates. This decoder is known as a NOR decoder. Observe that because of the precharge operation, the decoder circuit does not dissipate static power.

EXERCISE

18.12 How many transistors are needed for a NOR row decoder with an M-bit address?
Ans. $M2^M$ NMOS $+ 2^M$ PMOS $= 2^M(M+1)$

18.4.4 The Column-Address Decoder

From the description in Section 18.3, the function of the column-address decoder is to connect one of the 2^N bit lines to the data I/O line of the chip. As such, it is a multiplexer and can be implemented using pass-transistor logic (Section 18.1) as shown in Fig. 18.37. Here, each bit line is connected to the data I/O line through an NMOS transistor. The gates of the pass transistors are controlled by 2^N lines, one of which is selected by a NOR decoder similar to that used for decoding the row address. Finally, note that better performance can be obtained by utilizing transmission gates in place of NMOS transistors (Section 18.1). In such a case, however, the decoder needs to provide complementary output signals.

An alternative implementation of the column decoder that uses a smaller number of transistors (but at the expense of slower speed of operation) is shown in Fig. 18.38. This circuit, known as a *tree decoder,* has a simple structure of pass transistors. Unfortunately,

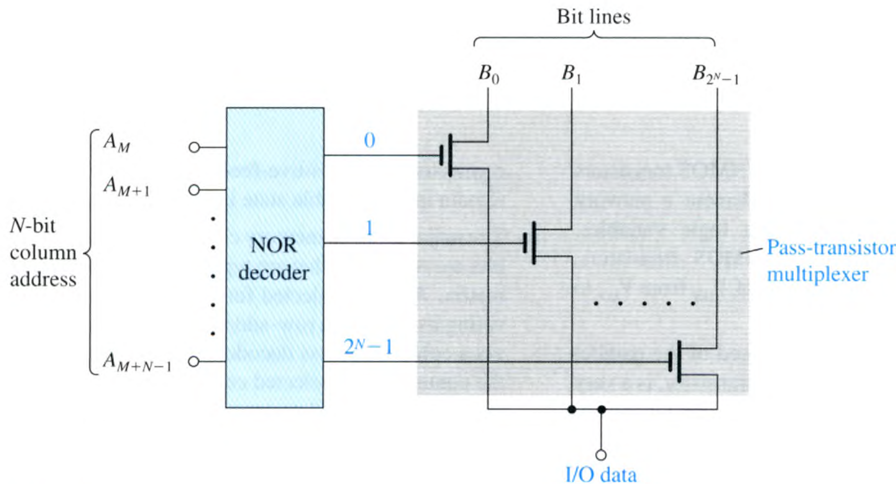


Figure 18.37 A column decoder realized by a combination of a NOR decoder and a pass-transistor multiplexer.

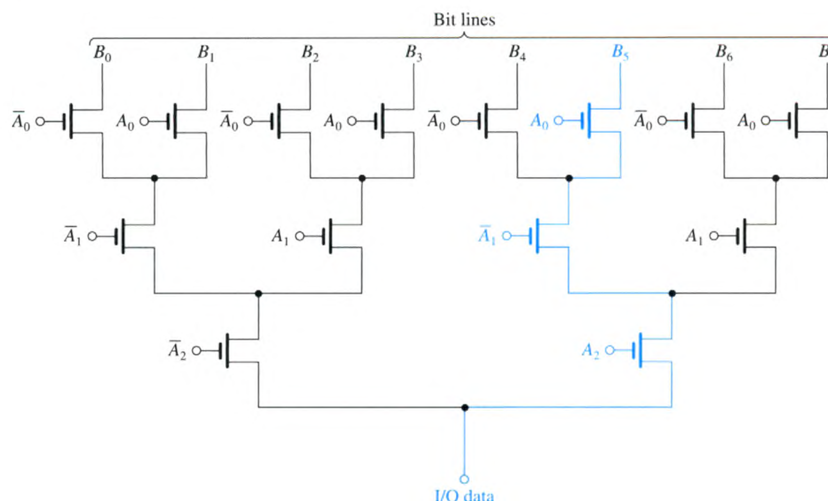


Figure 18.38 A tree column decoder. Note that the colored path shows the transistors that are conducting when $A_0 = 1$, $A_1 = 0$, and $A_2 = 1$, the address that results in connecting B_5 to the data line.

since a relatively large number of transistors can exist in the signal path, the resistance of the bit lines increases, and the speed decreases correspondingly.

EXERCISE

18.13 How many transistors are needed for a tree decoder when there are 2^N bit lines?
Ans. $2(2^N - 1)$

Summary

- Pass-transistor logic uses either single NMOS transistors or CMOS transmission gates to implement a network of switches controlled by the input logic variables. Switches implemented by single NMOS transistors, though simple, result in the reduction of V_{OH} from V_{DD} to $V_{DD} - V_t$.
- The CMOS transmission gate, composed of the parallel connection of an NMOS and a PMOS transistor, is a very effective switch in both analog and digital applications. It passes the entire input signal swing, 0 to V_{DD}. As well, it has an almost constant on-resistance over the full output range.
- Flip-flops employ one or more latches, often in a master-slave configuration. The basic static latch is a bistable circuit implemented using two inverters connected in a positive-feedback loop. The latch can remain in either stable state indefinitely.
- The major part of a memory chip consists of cells in which bits are stored and that are typically organized in a square matrix. A cell is selected for reading or writing by activating its row, via a row-address decoder, and its column, via a column-address decoder. A sense amplifier detects the content of the selected cell and provides a full-swing version of it to the data-output terminal of the chip.
- There are two kinds of MOS RAM: static and dynamic. Static RAMs (SRAMs) employ flip-flops as the storage cells. In a dynamic RAM (DRAM), data are stored on a capacitor and thus must be periodically refreshed. DRAM chips provide the highest possible storage capacity for a given chip area.

- Two circuits have emerged as the near-universal choice in implementing the storage cell: the six-transistor SRAM cell and the one-transistor DRAM cell.
- Proper read and write functioning of the 6T SRAM cell imposes constraints on the relative sizes of its transistors.
- Ring oscillators can be used to generate some timing signals in sequential systems. The frequency of these oscillators is determined by the number of inverters in the ring and the inverter propagation delay, which can be controlled by the supply voltage to create a voltage-controlled oscillator.
- Although sense amplifiers are used in SRAMs to speed up operation, they are essential in DRAMs. A particular type of sense amplifier is a differential circuit that employs positive feedback to obtain an output signal that grows exponentially toward either V_{DD} or 0.

PROBLEMS

Section 18.1: The Transmission Gate

18.1 Recall that MOS transistors are symmetrical and that what distinguishes the source from the drain is their relative voltage levels: For NMOS, the terminal with the higher voltage is the drain; for PMOS, the terminal with the higher voltage is the source. For each of the circuits in Fig. P18.1, label the source and drain terminals and give the output voltage V_O in terms of V_{DD}, V_{tn}, and $|V_{tp}|$. Note that V_{tn} and $|V_{tp}|$ are determined by the body effect, and give expressions for their values. Note that V_O is the value reached after the capacitor charging/discharging interval has come to an end.

18.2 Let the NMOS transistor switch in Fig. 18.3 be fabricated in a 0.13-μm CMOS process for which $V_{t0} = 0.4$ V, $\gamma = 0.2\ \text{V}^{1/2}$, $2\phi_f = 0.88$ V, and $V_{DD} = 1.2$ V. Determine V_{OH}.

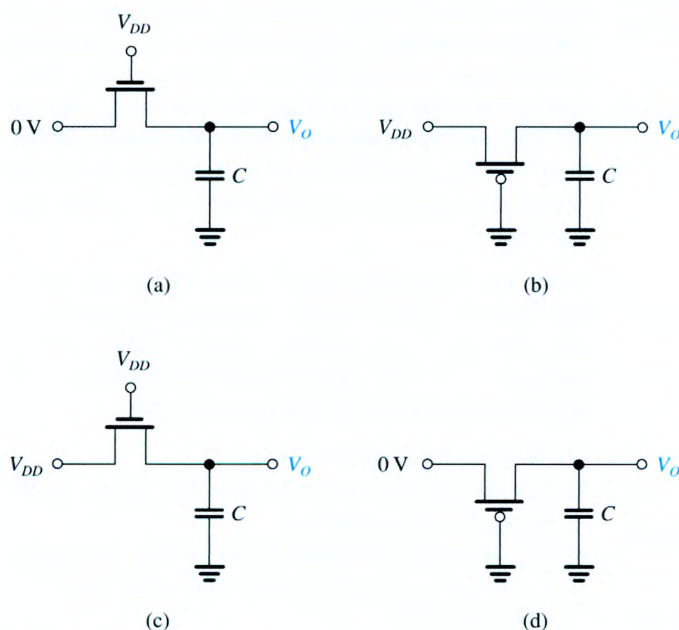


Figure P18.1

18.3 Consider the circuit in Fig. 18.3 with the NMOS transistor having $W/L = 1.5$ and fabricated in a CMOS process for which $V_{t0} = 0.4$ V, $\gamma = 0.2$ V$^{1/2}$, $2\phi_f = 0.88$ V, $V_{DD} = 1.2$ V, and $\mu_n C_{ox} = 500\ \mu$A/V^2. Find t_{PLH} for the case $C = 10$ fF.

18.4 Consider the circuit in Fig. 18.4 with the NMOS transistor having $W/L = 1.5$ and fabricated in a 0.13-μm CMOS process for which $V_{t0} = 0.4$ V, $V_{DD} = 1.2$ V, and $\mu_n C_{ox} = 500\ \mu$A/V^2. Determine t_{PHL} for the case $C = 10$ fF.

18.5 Consider the case specified in Exercise 18.1. If the output of the switch is connected to the input of a CMOS inverter having $(W/L)_p = 2(W/L)_n = 0.54$ μm/0.18 μm, find the static current of the inverter and its static power dissipation when the inverter input is at the value of V_{OH} found in Exercise 18.1. Also find the inverter output voltage. Let $\mu_n C_{ox} = 4\ \mu_p C_{ox} = 300\ \mu$A/V^2.

18.6 An NMOS pass-transistor switch with $W/L = 130$ nm/65 nm, used in a 1-V system for which $V_{t0} = 0.35$ V, $\gamma = 0.25$ V$^{1/2}$, $2\phi_f = 1.0$ V, $\mu_n C_{ox} = 3\mu_p C_{ox} = 450\ \mu$A/V^2, drives a 10-fF load capacitance at the input of a matched standard CMOS inverter using $(W/L)_n = 130$ nm/65 nm. For the switch gate terminal at V_{DD}, evaluate the switch V_{OH} and V_{OL} for inputs at V_{DD} and 0 V, respectively. For this value of V_{OH}, what inverter static current results? Estimate t_{PLH} and t_{PHL} for this arrangement as measured from the input to the output of the switch itself.

18.7 Figure P18.7 shows a PMOS transistor operating as a switch in the on position.

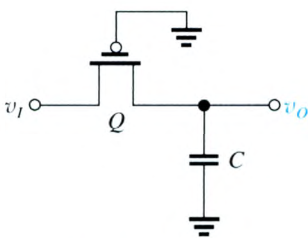


Figure P18.7

(a) If initially $v_O = 0$ and at $t = 0$, v_I is raised to V_{DD}, what is the final value V_{OH} reached at the output?

(b) If initially $v_O = V_{DD}$ and at $t = 0$, v_I is lowered to 0 V, what is the final value V_{OL} reached at the output?

(c) For the situation in (a), find t_{PLH} for v_O to rise from 0 to $V_{DD}/2$. Let $k_p = 125\ \mu$A/V^2, $V_{DD} = 1.2$ V, and $|V_{tp}| = 0.4$ V.

18.8 For the level-restoring circuit of Fig. 18.5, let $k_n' = 3k_p' = 450\ \mu$A/V^2, $V_{DD} = 1.0$ V, $|V_{t0}| = 0.35$ V, $\gamma = 0.5$ V$^{1/2}$, $2\phi_f = 1.0$ V, $(W/L)_1 = (W/L)_n = 130$ nm/65 nm, $(W/L)_p = 390$ nm/65 nm, and $C = 20$ fF. Also, let $v_B = V_{DD}$. Now, for v_A rising to V_{DD}, and Q_1 charging C and causing v_{O1} to rise, show that the value of v_{O1} that causes v_{O2} to drop by a threshold voltage below V_{DD} (i.e., to 0.65 V) so that Q_R turns on, is approximately $V_{DD}/2$ and thus occurs at $t \simeq t_{PLH}$. What is the capacitor-charging current available at this time (i.e., just prior to Q_R turning on)? What is it at $v_{O1} = 0$? What is the average current available for charging C? Estimate the time t_{PLH}. (Note that after Q_R turns on, v_{O1} rises to V_{DD}.)

D *18.9 The purpose of this problem is to illustrate how W/L of the level-restoring transistor Q_R in the circuit of Fig. 18.5 is determined. For this purpose consider the circuit as specified in Problem 18.8 and let $v_B = V_{DD}$. Now, consider the situation when v_A is brought down to 0 V and Q_1 conducts and begins to discharge C. The voltage v_{O1} will begin to drop from V_{DD}. Meanwhile, v_{O2} is still low and Q_R is conducting (though at $t = 0$, the current in Q_R is zero). Calculate the discharge current at $t = 0$. As Q_R conducts, its current subtracts from the current of Q_1, reducing the current available to discharge C. Find the value of v_{O1} at which the inverter begins to switch. This is $V_{IH} = \frac{1}{8}(5V_{DD} - 2V_t)$. Then, find the current that Q_1 conducts at this value of v_{O1}. Choose W/L for Q_R so that the maximum current it conducts is limited to one-half the value of the current in Q_1. What is the W/L you have chosen? Estimate t_{PHL} as the time for v_{O1} to drop from V_{DD} to V_{IH}.

18.10 The transmission gate in Figs. 18.7(a) and (b) is fabricated in a CMOS process technology for which $k_n' = 4k_p' = 500\ \mu$A/V^2, $|V_{t0}| = 0.4$ V, $\gamma = 0.2$ V$^{1/2}$, $2\phi_f = 0.88$ V, and $V_{DD} = 1.2$ V. Let Q_N and Q_P have $(W/L)_n = (W/L)_p = 1.5$. The total capacitance at the output node is 15 fF.

(a) What are the values of V_{OH} and V_{OL}?

(b) For the situation in Fig. 18.7(a), find $i_{DN}(0)$, $i_{DP}(0)$, $i_{DN}(t_{PLH})$, $i_{DP}(t_{PLH})$, and t_{PLH}.

(c) For the situation depicted in Fig. 18.7(b), find $i_{DN}(0)$, $i_{DP}(0)$, $i_{DN}(t_{PHL})$, $i_{DP}(t_{PHL})$, and t_{PHL}. At what value of v_O will Q_P turn off?

(d) Find t_P.

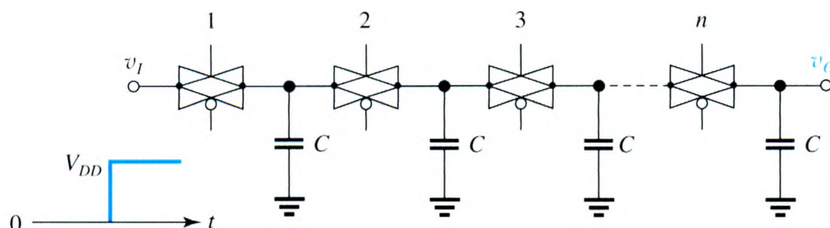


Figure P18.13

18.11 For the transmission gate specified in Problem 18.10, find R_{TG} at $v_O = 0$ and 0.6 V. Use the average of those values to determine t_{PLH} for the situation in which $C = 15$ fF.

***18.12** Refer to the situation in Fig. 18.7(b). Derive expressions for R_{Neq}, R_{Peq}, and R_{TG} following the approach used in Section 18.1.3 for the capacitor-charging case. Evaluate the value of R_{TG} for $v_O = V_{DD}$ and $v_O = V_{DD}/2$ for the process technology specified in Problem 18.10. Find the average value of R_{TG} and use it to determine t_{PHL} for the case $C =$ 15 fF.

18.13 Figure P18.13 shows a chain of transmission gates. This situation often occurs in circuits such as adders and multiplexers. Consider the case when all the transmission gates are turned on and a step voltage V_{DD} is applied to the input. The propagation delay t_P can be determined from the Elmore delay formula as follows:

$$t_P = 0.69\sum_{k=1}^{n} kCR_{TG}$$

where R_{TG} is the resistance of each transmission gate, C is the capacitance between each node and ground, and n is the number of transmission gates in the chain. Note that the sum of the series in this formula is given by

$$t_P = 0.69\,CR_{TG}\frac{n(n+1)}{2}$$

Now evaluate t_P for the case of 16 transmission gates with $R_{TG} = 10$ kΩ and $C = 10$ fF. What does the value of t_P become if the input is a ramp rather than a step function?

Section 18.2: Latches and Flip-Flops

18.14 Consider the latch of Fig. 18.11 with the two inverters identical and each characterized by $V_{OL} = 0$ V, $V_{OH} = 1.2$ V, $V_{IL} = 0.5$ V, and $V_{IH} = 0.7$ V. Let the transfer characteristic of each inverter be approximated by three straight-line segments. Sketch the transfer characteristic of the feedback loop of the latch and give the coordinates of points A, B, and C (refer to Fig. 18.11). What is the gain at point B? What is the width of the transition region?

D 18.15 Sketch the standard CMOS circuit implementation of the SR flip-flop shown in Fig. 18.13.

D 18.16 Sketch the logic-gate implementation of an SR flip-flop utilizing two cross-coupled NAND gates. Clearly label the output terminals and the input trigger terminals. Provide the truth table and describe the operation.

D 18.17 For the SR flip-flop of Fig. 18.14, show that if each of the two inverters utilizes matched transistors, that is, $(W/L)_p = (\mu_n/\mu_p)(W/L)_n$, then the minimum W/L that each of Q_5–Q_8 must have so that switching occurs is $2(W/L)_n$. Give the sizes of all eight transistors if the flip-flop is fabricated in a 65-nm process for which $\mu_n = 3\mu_p$. Use the minimum channel length for all transistors and the minimum size ($W/L = 1$) for Q_1 and Q_3.

D 18.18 Repeat part (a) of the problem in Example 18.2 for the case of inverters that do not use matched Q_N and Q_P. Rather, assume that each of the inverters uses $(W/L)_p = (W/L)_n = 130$ nm/65 nm. Find the threshold voltage of each inverter. Then determine the value required for the W/L of each of Q_5 to Q_8 so that the flip-flop switches. (*Hint:* Refer to Table 16.2.)

D 18.19 The CMOS SR flip-flop in Fig. 18.14 is fabricated in a 0.13-μm process for which $\mu_n C_{ox} = 4\mu_p C_{ox} =$ 500 μA/V^2, $V_{tn} = |V_{tp}| = 0.4$ V, and $V_{DD} = 1.2$ V. The inverters have $(W/L)_n = 0.2$ μm/0.13 μm and $(W/L)_p = 0.8$ μm/0.13 μm. The four NMOS transistors in the set–reset circuit have equal W/L ratios.

(a) Determine the minimum value required for this ratio to ensure that the flip-flop will switch.

(b) If a ratio twice the minimum is selected, determine the minimum required width of the set and reset pulses to ensure switching. Assume that the total capacitance between each of the Q and $\overline{Q}$ nodes and ground is 15 fF.

D 18.20 The clocked SR flip-flop in Fig. 18.14 is not a fully complementary CMOS circuit. Sketch the fully complementary version by augmenting the circuit with the PUN corresponding to the PDN comprising Q_5, Q_6, Q_7, and Q_8. Note that the fully complementary circuit utilizes 12 transistors. Although the circuit is more complex, it switches faster.

D 18.21 Consider another possibility for the circuit in Fig. 18.17: Relabel the R input as $\overline{S}$ and the S input as $\overline{R}$. Let $\overline{S}$ and $\overline{R}$ normally rest at V_{DD}. Let the flip-flop be storing a 0; thus $V_Q = 0$ V and $V_{\overline{Q}} = V_{DD}$. To set the flip-flop, the $\overline{S}$ terminal is lowered to 0 V and the clock ϕ is raised to V_{DD}. The relevant part of the circuit is then transistors Q_5 and Q_2. For the flip-flop to switch, the voltage at $\overline{Q}$ must be lowered to $V_{DD}/2$. What is the minimum required W/L for Q_5 in terms of $(W/L)_2$ and (μ_n/μ_p)? Assume $V_{tn} = |V_{tp}|$.

***18.22** Figure P18.22 shows a commonly used circuit of a D flip-flop that is triggered by the negative-going edge of the clock ϕ.

(a) For ϕ high, what are the values of $\overline{Q}$ and Q in terms of D? Which transistors are conducting?
(b) If D is high and ϕ goes low, which transistors conduct and what signals appear at $\overline{Q}$ and at Q? Describe the circuit operation.
(c) Repeat (b) for D low with the clock ϕ going low.
(d) Does the operation of this circuit rely on charge storage?

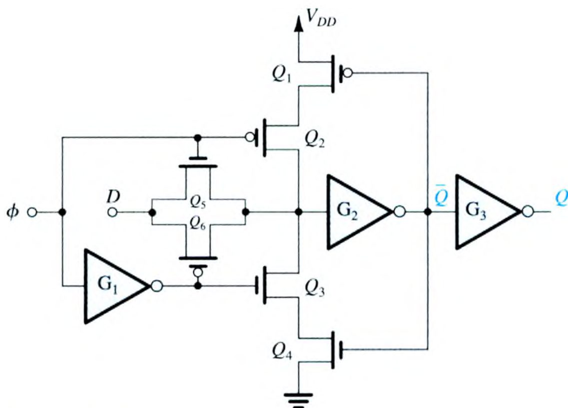


Figure P18.22

Section 18.3: Random-Access Memory (RAM) Cells

18.23 Repeat Exercise 18.6 for an SRAM fabricated in a 0.13-μm CMOS process for which $V_{DD} = 1.2$ V and $V_t = 0.4$ V. Use $(W/L)_n = 1.5$.

18.24 Locate on the graph of Fig. 18.24 the points (a), (b), and (c) that correspond to the following three process technologies:

(a) 0.18-μm: $V_{DD} = 1.8$ V and $V_t = 0.5$ V
(b) 0.13-μm: $V_{DD} = 1.2$ V and $V_t = 0.4$ V
(c) 65-nm: $V_{DD} = 1.0$ V and $V_t = 0.35$ V

In each case, impose the condition that in a read-1 operation $V_{\overline{Q}} = V_t$.

D 18.25 Find the maximum allowable W/L for the access transistors of the SRAM cell in Fig. 18.22 so that in the read operation, the voltages at Q and $\overline{Q}$ do not change by more than $|V_t|$. Assume that the SRAM is fabricated in a 0.13-μm technology for which $V_{DD} = 1.2$ V and $V_{tn} = |V_{tp}| = 0.4$ V, and $(W/L)_n = 1.5$. Find $V_{\overline{Q}}$ and I_5 that result in each of the following cases:

(a) $(W/L)_a = \frac{1}{3}$ the maximum allowed
(b) $(W/L)_a = \frac{2}{3}$ the maximum allowed
(c) $(W/L)_a =$ the maximum allowed

Assume $\mu_n C_{ox} = 500$ μA/V^2. Which one of the three designs results in the shortest read delay?

D 18.26 Consider a 6T SRAM cell fabricated in a 65-nm CMOS process for which $V_{tn} = |V_{tp}| = 0.35$ V and $V_{DD} = 1.0$ V. If during a read-1 operation it is required that $V_{\overline{Q}}$ not exceed 0.1 V, use the graph in Fig. 18.24 to determine the maximum allowable value of the ratio $(W/L)_5/(W/L)_1$. For $L_1 = L_5 = 65$ nm, select values for W_1 and W_5 that minimize the combined areas of Q_1 and Q_5. Assume that the minimum width allowed is 65 nm.

18.27 Consider the read operation of the 6T SRAM cell of Fig. 18.22 when it is storing a 0, that is, $V_Q = 0$ V, and $V_{\overline{Q}} = V_{DD}$. Assume that the bit lines are precharged to V_{DD} before the word-line voltage is raised to V_{DD}. Sketch the relevant part of the circuit and describe the operation. Show that the analysis parallels that presented in the text for the read-1 operation.

D *18.28 For the 6T SRAM of Fig. 18.22, fabricated in a 0.13-μm CMOS process for which $V_{DD} = 1.2$ V, $V_{t0} = 0.4$ V, $2\phi_f = 0.88$ V, and $\gamma = 0.2\ \text{V}^{1/2}$, find the maximum ratio $(W/L)_5/(W/L)_1$ for which $V_{\overline{Q}} \leq V_{t0}$ during a read-1 operation (Fig. 18.23). Then, take into account the body effect in Q_5 and compare this result to the value obtained without accounting for the body effect.

D 18.29 A 6T SRAM cell is fabricated in a 0.13-μm CMOS process for which $V_{DD} = 1.2$ V, $V_t = 0.4$ V, and $\mu_n C_{ox} = 500\ \mu\text{A/V}^2$. The inverters utilize $(W/L)_n = 1$. Each of the bit lines has a 2-pF capacitance to ground. The sense amplifier requires a minimum of 0.2-V input for reliable and fast operation.

(a) Find the upper bound on W/L for each of the access transistors so that V_Q and $V_{\overline{Q}}$ do not change by more than V_t volts during the read operation.
(b) Find the delay time Δt encountered in the read operation if the cell design utilizes minimum-size access transistors.
(c) Find the delay time Δt if the design utilizes the maximum allowable size for the access transistors.

18.30 Consider the operation of writing a 1 into a 6T SRAM cell that is originally storing a 0. Sketch the relevant part of the circuit and explain the operation. Without doing detailed analysis, show that the analysis would lead to results identical to those obtained in the text for the write-0 operation.

D 18.31 For a 6T SRAM cell fabricated in a 0.13-μm CMOS process, find the maximum permitted value of $(W/L)_p$ in terms of $(W/L)_a$ of the access transistors. Assume $V_{DD} = 1.2$ V, $V_{tn} = |V_{tp}| = 0.4$ V, and $\mu_n = 4\mu_p$.

18.32 Locate on the graph in Fig. 18.27 the points (a), (b), and (c) corresponding to the following three CMOS fabrication processes:

(a) 0.18-μm: $V_{DD} = 1.8$ V, $V_{tn} = |V_{tp}| = 0.5$ V
(b) 0.13-μm: $V_{DD} = 1.2$ V, $V_{tn} = |V_{tp}| = 0.4$ V
(c) 65-nm: $V_{DD} = 1.0$ V, $V_{tn} = |V_{tp}| = 0.35$ V

For all three, $\mu_n \simeq 4\mu_p$. In each case, V_Q is to be limited to a maximum value of V_{tn}.

D 18.33 Design a minimum-size 6T SRAM cell in a 0.13-μm process for which $V_{DD} = 1.2$ V and $V_{tn} = |V_{tp}| = 0.4$ V. All transistors are to have equal $L = 0.13$ μm. Assume that the minimum width allowed is 0.13 μm. Verify that your minimum-size cell meets the constraints in Eqs. (18.23) and (18.29).

18.34 For a particular DRAM design, the cell capacitance $C_S = 35$ fF and $V_{DD} = 1.2$ V. Each cell represents a capacitive load on the bit line of 0.8 fF. Assume a 20-fF capacitance for the sense amplifier and other circuitry attached to the bit line. What is the maximum number of cells that can be attached to a bit line while ensuring a minimum bit-line signal of 25 mV? How many bits of row addressing can be used? If the sense-amplifier gain is increased by a factor of 4, how many word-line address bits can be accommodated?

18.35 In a particular dynamic memory chip, $C_S = 30$ fF, the bit-line capacitance per cell is 0.5 fF, and bit-line control circuitry involves 12 fF. For a 1-Mbit-square array, what bit-line signals result when a stored 1 is read? When a stored 0 is read? Assume that $V_{DD} = 1.0$ V.

18.36 For a DRAM cell utilizing a capacitance of 25 fF, refresh is required within 12 ms. If a signal loss on the capacitor of 0.2 V can be tolerated, what is the largest acceptable leakage current present at the cell?

Section 18.4: Ring Oscillators and Special-Purpose Circuits

18.37 Consider a ring oscillator consisting of five inverters, each having $t_{PLH} = 300$ ps and $t_{PHL} = 200$ ps. Sketch one of the output waveforms, and specify its frequency and the percentage of the cycle during which the output is high.

18.38 A ring-of-nine oscillator is found to operate at 250 MHz. Find the propagation delay of the inverter. If the supply voltage is reduced by 20%, at what frequency do you expect the circuit to oscillate?

D 18.39 Design the one-shot circuit of Fig. 18.33 to provide an output pulse of 0.8-ns width. If the inverters available have $t_P = 200$-ps delay, how many inverters do you need for the delay circuit?

D 18.40 Consider the operation of the differential sense amplifier of Fig. 18.34 following the rise of the sense control signal ϕ_s. Assume that a balanced differential signal of 0.1 V is established between the bit lines, each of which has a 1 pF capacitance. For $V_{DD} = 1.2$ V, what value of G_m of each of the inverters in the amplifier is required to cause the outputs to reach $0.1V_{DD}$ and $0.9V_{DD}$ [from initial values of $0.5V_{DD} - (0.1/2)$ and $0.5V_{DD} + (0.1/2)$ volts, respectively] in 2 ns? If for the matched inverters, $|V_t| = 0.4$ V and $k_n' = 4k_p' = 500\ \mu\text{A/V}^2$, what are the device widths

required? If the input signal is 0.2 V, what does the amplifier response time become?

18.41 A particular version of the regenerative sense amplifier of Fig. 18.34 in a 0.13-μm technology uses transistors for which $|V_t| = 0.4$ V, $k_n' = 4k_p' = 500\ \mu A/V^2$, $V_{DD} = 1.2$ V, with $(W/L)_n = 0.26\ \mu m/0.13\ \mu m$ and $(W/L)_p = 1.04\ \mu m/0.13\ \mu m$. For each inverter, find the value of G_m. For a bit-line capacitance of 0.4 pF, and a delay until an output of $0.9V_{DD}$ is reached of 1 ns, find the initial difference voltage required between the two bit lines. If the time can be relaxed by 1 ns, what input signal can be handled? With the increased delay time and with the input signal at the original level, by what percentage can the bit-line capacitance, and correspondingly the bit-line length, be increased? If the delay time required for the bit-line capacitances to charge by the constant current available from the storage cell, and thus develop the difference-voltage signal needed by the sense amplifier, was 2 ns, what does it increase to when longer lines are used?

D 18.42 (a) For the sense amplifier of Fig. 18.34, show that the time required for the bit lines to reach $0.9V_{DD}$ and $0.1V_{DD}$ is given by $t_d = (C_B/G_m)\ln(0.8V_{DD}/\Delta V)$, where ΔV is the initial difference voltage between the two bit lines.

(b) If the response time of the sense amplifier is to be reduced to one-half the value of an original design, by what factor must the width of all transistors be increased?

(c) If for a particular design, $V_{DD} = 1.2$ V and $\Delta V = 0.2$ V, find the factor by which the widths of all transistors must be increased so that ΔV is reduced by a factor of 2, while keeping t_d unchanged?

18.43 Consider a 1024-row NOR decoder. To how many address bits does this correspond? How many output lines does the decoder have? How many input lines does the NOR array require? How many NMOS and PMOS transistors does such a design need?

18.44 For the column decoder shown in Fig. 18.37, how many column-address bits are needed in a 1-Mbit-square array? How many NMOS pass transistors are needed in the multiplexer? How many NMOS transistors are needed in the NOR decoder? How many PMOS transistors? What is the total number of NMOS and PMOS transistors needed?

18.45 Consider the use of the tree column decoder shown in Fig. 18.37 for application with a square 1-Mbit array. How many address bits are involved? How many levels of pass gates are used? How many pass transistors are there in total?

APPENDICES ON COMPANION WEBSITE

For your convenience, ten additional chapters on important reference topics are included on the companion website. In PDF format, the appendices are fully searchable and can be bookmarked.

Appendix A: VLSI Fabrication Technology This article is a concise explanation of the technology that goes into fabricating integrated circuits. The different processes used are described and compared, and the characteristics of the resulting devices presented. Design considerations that restrict IC designers are explored.

Appendix B: SPICE Device Models and Design with Simulation Examples This appendix could stand as a book on its own. First, it describes the models SPICE programs use to represent op amps, diodes, MOSFETs, and BJTs in integrated circuits. A thorough understanding of these models is critical for designers trying to extract meaningful information from an analysis. Then, it describes and discusses all the simulations with reference to the corresponding netlists, also available on the companion website. This is a rich resource to help analyze, experiment with, and design circuits that relate to the topics studied in *Microelectronic Circuits.*

Appendix C: Two-Port Network Parameters Throughout the text, we use different possible ways to characterize linear two-port networks. This appendix summarizes the *y*, *z*, *h*, and *g* parameters and provides their equivalent-circuit representations.

Appendix D: Some Useful Network Theorems This article reviews Thévenin's theorem, Norton's theorem, and the source-absorption theorem, all of which are useful in simplifying the analysis of electronic circuits.

Appendix E: Single-Time-Constant Circuits STC circuits are composed of, or can be reduced to, one reactive component (inductance or capacitance) and one resistance. This is important to the design and analysis of linear and digital circuits. Analyzing an amplifier circuit can usually be reduced to the analysis of one or more STC circuits.

Appendix F: *s*-Domain Analysis: Poles, Zeros, and Bode Plots Most of the work in analyzing the frequency response of an amplifier involves finding the amplifier voltage gain as a function of the complex frequency *s*. The tools to do this are summarized in this appendix.

Appendix G: Comparison of the MOSFET and the BJT Provides a comprehensive compilation and comparison of the properties of the MOSFET and the BJT. The comparison is aided by the inclusion of typical parameter values of devices fabricated with modern process technologies.

Appendix H: Filter Design Material This section features tables and graphs useful for the design of the filter circuits studied in Chapter 14.

Appendix I: Bibliography An excellent resource for students beginning research projects, this bibliography outlines key reference works on electronic circuits, circuit and system analysis, devices and IC fabrication, op amps, analog and digital circuits, filters and tuned amplifiers, and SPICE.

Appendix L: Answers to Selected Problems.

APPENDIX J

STANDARD RESISTANCE VALUES AND UNIT PREFIXES

Discrete resistors are available only in standard values. Table J.1 provides the multipliers for the standard values of 5%-tolerance and 1%-tolerance resistors. Thus, in the kilohm range of 5% resistors, one finds resistances of 1.0, 1.1, 1.2, 1.3, 1.5,.... In the same range, one finds 1% resistors of kilohm values of 1.00, 1.02, 1.05, 1.07, 1.10,....

Table J.1 Standard Resistance Values

	1% Resistor Values (kΩ)			
5% Resistor Values (kΩ)	100–174	178–309	316–549	562–976
10	100	178	316	562
11	102	182	324	576
12	105	187	332	590
13	107	191	340	604
15	110	196	348	619
16	113	200	357	634
18	115	205	365	649
20	118	210	374	665
22	121	215	383	681
24	124	221	392	698
27	127	226	402	715
30	130	232	412	732
33	133	237	422	750
36	137	243	432	768
39	140	249	442	787
43	143	255	453	806
47	147	261	464	825
51	150	267	475	845
56	154	274	487	866
62	158	280	499	887
68	162	287	511	909
75	165	294	523	931
82	169	301	536	953
91	174	309	549	976

Table J.2 provides the SI unit prefixes used in this book and in all modern works in English.

Table J.2 SI Unit Prefixes

Name	Symbol	Factor
femto	f	$\times 10^{-15}$
pico	p	$\times 10^{-12}$
nano	n	$\times 10^{-9}$
micro	μ	$\times 10^{-6}$
milli	m	$\times 10^{-3}$
kilo	k	$\times 10^{3}$
mega	M	$\times 10^{6}$
giga	G	$\times 10^{9}$
tera	T	$\times 10^{12}$
peta	P	$\times 10^{15}$

Table J.3 provides the meter conversion factors.

Table J.3 Meter Conversion Factors

$1\ \mu m = 10^{-4}\ cm = 10^{-6}\ m$
$1\ m = 10^{2}\ cm = 10^{6}\ \mu m = 10^{9}\ nm$
$0.1\ \mu m = 100\ nm$
$1\ \text{Å} = 10^{-8}\ cm = 10^{-10}\ m$

APPENDIX K

TYPICAL PARAMETER VALUES FOR IC DEVICES FABRICATED IN CMOS AND BIPOLAR PROCESSES

Table K.1 Typical Values of CMOS Device Parameters

	0.8 μm		0.5 μm		0.25 μm		0.18 μm		0.13 μm		65 nm		28 nm	
Parameter	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS
t_{ox} (nm)	15	15	9	9	6	6	4	4	2.7	2.7	1.4	1.4	1.4	1.4
C_{ox} (fF/μm^2)	2.3	2.3	3.8	3.8	5.8	5.8	8.6	8.6	12.8	12.8	25	25	34	34
μ (cm^2/V·s)	550	250	500	180	460	160	450	100	400	100	216	40	220	200
μC_{ox} (μA/V^2)	127	58	190	68	267	93	387	86	511	128	540	100	750	680
V_{t0} (V)	0.7	−0.7	0.7	−0.8	0.5	−0.6	0.5	−0.5	0.4	−0.4	0.35	−0.35	0.3	−0.3
V_{DD} (V)	5	5	3.3	3.3	2.5	2.5	1.8	1.8	1.3	1.3	1.0	1.0	0.9	0.9
$\lvert V_A' \rvert$ (V/μm)	25	20	20	10	5	6	5	6	5	6	3	3	1.5	1.5
C_{ov} (fF/μm)	0.2	0.2	0.4	0.4	0.3	0.3	0.37	0.33	0.36	0.33	0.33	0.31	0.4	0.4

Table K.2 Typical Parameter Values for BJTs*

	Standard High-Voltage Process		Advanced Low-Voltage Process	
Parameter	*npn*	Lateral *pnp*	*npn*	Lateral *pnp*
A_E (μm^2)	500	900	2	2
I_S (A)	5×10^{-15}	2×10^{-15}	6×10^{-18}	6×10^{-18}
β_0 (A/A)	200	50	100	50
V_A (V)	130	50	35	30
V_{CEO} (V)	50	60	8	18
τ_F	0.35 ns	30 ns	10 ps	650 ps
C_{je0}	1 pF	0.3 pF	5 fF	14 fF
$C_{\mu 0}$	0.3 pF	1 pF	5 fF	15 fF
r_x (Ω)	200	300	400	200

*Adapted from Gray et al. (2001); see Appendix I.

Note: For more information, refer to Appendix G (on the book's Website).

INDEX

References marked **P** are study Problems; **S** are points in the chapter Summaries; Appendix pages found in the book are shown as J1

C

D

G

H

N

O

P

T

U

V

W

X

Z



BINDER-READY EDITION: Flexible format at an even more affordable price

SEDRA/SMITH, educating over a million microelectronic circuit students

Microelectronic Circuits by Sedra and Smith has served generations of electrical and computer engineering students as the best and most widely-used text for this required course. Respected equally as a textbook and reference, "Sedra/Smith" combines a thorough presentation of fundamentals with an introduction to present-day IC technology. It remains the best text for helping students progress from circuit analysis to circuit design, developing design skills and insights that are essential to successful practice in the field. Significantly revised with the input of two new coauthors, slimmed down, and updated with the latest innovations, *Microelectronic Circuits,* Eighth Edition, remains the gold standard in providing the most comprehensive, flexible, accurate, and design-oriented treatment of electronic circuits available today.

New to this edition

- **New digital features**: Forty step-by-step problem-solving videos—which help students bridge the gap from theory to application—are embedded in the eBook and accessible on a premium site for print purchasers.
- **Two new coauthors**—Tony Chan Carusone of University of Toronto and Vincent Gaudet of University of Waterloo—add a fresh perspective rooted in their active teaching and research.
- **Thoroughly updated for increased modularity and flexibility:** Changes have been made throughout most chapters to make it easier to teach if skipping chapters or covering them in a different order.
- **Overhaul of problem set (half of 1400 new or revised):** As IC technology is constantly changing, problem statements have all been updated, as needed, to ensure they reflect the parameters of newer generations of IC technologies.
- **Streamlined coverage (200 page reduction):** The authors place a greater emphasis on essentials to make it easier for students to grasp fundamentals and apply them to today's IC design challenges.

About the Authors

Adel S. Sedra is Distinguished Professor Emeritus of Electrical and Computer Engineering at the University of Waterloo and Distinguished Fellow, University Leadership, at Ryerson University.

Kenneth C. (KC) Smith is Professor Emeritus in Electrical and Computer Engineering, Computer Science, Industrial and Mechanical Engineering, and Information Studies at the University of Toronto.

Tony Chan Carusone is Professor of Electrical and Computer Engineering at the University of Toronto.

Vincent Gaudet is Professor and Chair in the Department of Electrical and Computer Engineering at the University of Waterloo.

OXFORD
UNIVERSITY PRESS
www.oup.com/us/he

ISBN 978-0-19-085354-9



Cover Design: M. Laseau